

VISIT...

LANZAROTE
Caliente.COM

KIRK-OTHMER

ENCYCLOPEDIA OF

CHEMICAL

TECHNOLOGY

EDITING - by **WATCHER**

VOLUME 21

FOURTH EDITION

RECYCLING TO
SILICON AND SILICON ALLOYS

CONTENT INDEX (vol 21)

(with hyperlinks)

Edited by <http://www.watcherworld.narod.ru/>

(RECYCLING)	
- OIL	1
- PAPER	6
- RUBBER	12
REFRACTORIES	22
REFRACTORY COATINGS	40
REFRACTORY FIBERS	53
REFRIGERATION	59
REGULATORY AGENCIES	72
- SURVEY	73
- CHEMICAL PROCESS INDUSTRY	76
- PHARMACEUTICALS AND COSMETICS	83
- POWER GENERATION	88
REINFORCED PLASTICS	94
RELEASE AGENTS	99
RENEWABLE ENERGY RESOURCES	104
REPELLENTS	112
RESEARCH/TECHNOLOGY MANAGEMENT	125
RESINS, NATURAL	138
REVERSE OSMOSIS	144
RHENIUM AND RHENIUM COMPOUNDS	160
RHEOLOGICAL MEASUREMENTS	166
ROOFING MATERIALS	209
RUBBER CHEMICALS	219
RUBBER COMPOUNDING	230
RUBBER, NATURAL	265
RUBIDIUM AND RUBIDIUM COMPOUNDS	278
SALICYLIC ACID AND RELATED COMPOUNDS	283
SAMPLING	296
SEALANTS	308
SEDIMENTATION	316
SELENIUM AND SELENIUM COMPOUNDS	326
SEMICONDUCTORS	342
- SILICON-BASED SEMICONDUCTORS	343
- AMORPHOUS SEMICONDUCTORS	357
- COMPOUND SEMICONDUCTORS	365
SENSORS	389
SEPARATION	395
- CENTRIFUGAL SEPARATION	396
- MAGNETIC SEPARATION	419
- SIZE SEPARATION	433
SEPARATIONS PROCESS SYNTHESIS	444
SHAPE-MEMORY ALLOYS	461
SILICA	468
- INTRODUCTION	469
- AMORPHOUS SILICA	483
- VITREOUS SILICA	497
- SYNTHETIC QUARTZ CRYSTALS	518
SILICON AND SILICON ALLOYS	523
- PURE SILICON	524
- CHEMICAL AND METALLURGICAL	535

OIL

The term oil includes a variety of liquid or easily liquefiable, unctuous, combustible substances that are soluble in ether but not in water and that leave a greasy stain on paper and cloth. These substances can include animal, vegetable, and synthetic oils, but usually the word oil refers to a mineral oil produced from petroleum (qv). An oil that has been used or contaminated, or both, but not consumed, can often be recycled to regain a useful material, regardless of its origin. For the purposes of this article, only the recycling of used petroleum oils is considered.

In the past, used or contaminated oil was usually considered a waste and was disposed of in a variety of ways, including incineration, road oiling, landspreading, municipal landfills, and dumping on the ground or into storm sewers. However, on account of the increased concern about pollution, the advent of periodic petroleum shortages, and increased prices for most kinds of oils and oil products, there is a strong interest in developing ways to conserve the valuable energy and resource content of these products.

The following definitions are useful to the discussion of oil recycling (1).

Used Oil.

Oil whose characteristics have changed since original manufacture and which is suitable for recycling. This is an umbrella category that includes used lubricating oils of all types as well as dirty or contaminated fuel or other oils that can be economically recycled.

Waste Oil.

Oil having characteristics that make it unsuitable for further use or economic recycling. This material may be usable as a fuel in large industrial furnaces, such as cement kilns.

Oil Recycling.

The acquisition and processing of oil that has become unsuitable for its intended use in order to regain useful material.

Oil Reclaiming or Oil Laundering.

The use of cleaning methods during recycling, primarily to remove insoluble contaminants, making the oil suitable for zfurther use. Methods may include settling, heating, dehydration, filtration, and centrifuging. The product is reclaimed oil.

Oil Re-Refining.

The use of refining processes during recycling to produce high quality lubricating base stock for lubricants or other petroleum products. Re-refining may include distillation, hydrotreating, and treatments employing acid, caustic, solvent, clay, and/or other chemicals. The product is re-refined lubricating base stock.

Oil Reprocessing.

The preparation of used oil for application as a fuel. Reprocessing may involve settling, filtration, and blending. This term has not been defined by ASTM as of this writing (1996).

Oil recycling dates to ca 1920. At that time, lubricating oils contained few or no additives, ie, chemical compounds added to oils to improve lubrication characteristics such as wear, oxidation, and corrosion (see LUBRICATION AND LUBRICANTS). Recycling these oils required only limited processing involving some combination of heating to remove volatile components; settling to separate water, dirt, and sludge; and centrifuging or filtering to remove most of the remaining insoluble contaminants. By this limited processing, lubricating oils could be recycled to essentially original oil quality. Since 1960, however, the types of lubricating oils in the marketplace have expanded and many contain high levels (up to 20% or more by volume) of complex additives, which can make recycling a much more difficult task (2).

Lubricating oil sales in the United States since the 1970s have remained relatively steady at $9.5 \times 10^6 \text{ m}^3/\text{yr}$ ($2.5 \times 10^9 \text{ gal/yr}$). Of this, automotive lubricants represent 47% and industrial lubricants take up the remainder (3,4). These industrial lubricants include the following types: hydraulic, quenching, cutting, metalworking, electrical, and process oils. From the $9.5 \times 10^6 \text{ m}^3$ of lubricating oils sold in the United States, ca $4.8 \times 10^6 \text{ m}^3$ ($1.3 \times 10^9 \text{ gal}$) of collectible used oil are generated annually. It is estimated that, of the 70% actually collected each year, ~7% is re-refined into motor oil, ~17% is reclaimed into industrial oil, and ~50% is burned as a fuel oil or used in road oiling and dust control, often undergoing little processing to remove contaminants (5). The unidentified portion (~25%) of the oil collected, along with the large quantity not collected, illustrates the significant pollution potential represented by improperly managed used oil.

There are ca 200 oil recyclers in the United States in the 1990s. Of these, only two are primarily re-refiners, which take the oil to be recycled through the entire refining process. The other companies involved in oil recycling utilize other, less drastic processes. Recyclers fall into three categories: reprocessors who process used oil for reuse as a fuel, often only removing free-standing water and larger pieces of dirt or rust; reclaimers who recover industrial oils from segregated oils, usually for reuse in the oils' original application; and re-refiners who produce base oil for blending a wide variety of lubricants. In many cases, an oil recycler is involved with more than one of these categories.

Recycled lubricating oil products are potentially suitable for all uses, including their original use, if given proper segregation, cleanup, and additive treatment, because the basic hydrocarbon structure is not significantly altered during most uses (2,6). Although nontechnical factors such as economics and availability determine whether all potential uses for recycled oil can be made available on a realistic basis, energy and natural resource conservation, pollution control, and improved international balance of payments all suggest that a strong effort be made in effective oil recycling.

Characteristics of Used Oils

Used petroleum oils to be recycled can be obtained from a variety of sources, including automotive garages and service stations, truck and taxi fleets, military installations, individuals, industrial plants, manufacturing facilities, and wastewater treatment plants. The main types of used petroleum oils that are recycled are internal combustion engine lubricants and hydraulic and industrial oils. The additives and contaminants typical in these oils can cause both performance-related and environmental problems (5,7,8). Much of the easily recoverable used oil in the 1990s, which has undergone little or no processing for contaminant removal, is employed as fuel (3,4). Used oils may often be commingled with each other and with water, solvents, and other chemicals before being collected for recycling.

Chemical analyses have been performed on used oils and waste oils, primarily for inorganic constituents. These analyses have been described in the literature (3,7,9–18) and are summarized in Table 1. The data show that used oils contain contaminants that often should be removed during the recycling process to protect the environment. However, many organic contaminants contained in such oils are not often analyzed. In particular, polychlorinated biphenyls (PCBs), which are strictly regulated at concentrations as low as 50 ppm (0.005 wt %) in some oils and at 2 ppm (0.0002 wt %) for other applications, are often not determined in used or waste oils, even though they sometimes occur in these oils (3,4) (see CHLOROCARBONS AND CHLOROHYDROCARBONS, TOXIC AROMATICS). These regulations on PCBs are established by the Toxic Substances Control Act (19); additional regulations are based on hazardous waste provisions in the Resource Conservation and Recovery Act (20). Chlorinated solvents are found frequently, usually as a result of contamination of the used oil during storage prior to recycling (8).

Table 1. Summary of Reported Used Oil Analyses^a

Property or test	Motor oils	Industrial oils ^b
viscosity, at 40°C, SUs	87–837	143–330
API gravity, at 15.6°C	19.1–31.3	25.7–26.2
specific gravity, 15.6°C/15.6°C	0.9396–0.8692	0.9002–0.8972
water, vol %	0.2–33.8 ^c	0.1–95
bottom sediment and water, vol %	0.1–42	
benzene insoluble, wt %	0.56–3.33	
gasoline dilution, vol %	2.0–9.7	
flash point, °C	79–220	157–179
heating value, MJ/kg ^d	31.56–44.88	40.12–41.84
ash, sulfated, wt %	0.03–6.43	3.2–5.9 ^e
carbon residue, wt %	1.82–4.43	
fatty oil, wt %		0–60
chlorine, wt %	0.17–0.47	<0.1–0.83
sulfur, wt %	0.17–1.09	0.54–1.03
zinc, ppm	260–1,787	
calcium, ppm	211–2,291	
barium, ppm	9–3,906	
phosphorus, ppm	319–1,550	
lead, ppm	85–21,676	
aluminum, ppm	<0.5–758	
iron, ppm	97–2,401	

^a Data taken from Refs. 3, 7, 9–18 did not provide data on all tests listed; therefore, data may be inconsistent between different tests.

^b Limited data were available for used industrial oils; see Ref. 6.

^c One sample had a water content of 46.5 wt % but is considered an outlier.

^d To convert MJ /kg to Btu/lb, multiply by 430.4.

^e Values for the industrial oils were stated to be for the regular, not the sulfated ash.

More recent data were published by the U.S. EPA in 1991. The summary in Table 2 shows the effect of unleaded gasoline on lead levels in used oil. In addition, the EPA study focused more on specific chlorinated solvents and polynuclear aromatics (PNAs). These compounds are considered when deciding whether a used oil must be handled as a hazardous waste (22). Some contaminants, such as chlorinated solvents, are picked up by the used oil after application and during storage waiting for collection. However, not all chlorine found in used oils is necessarily the result of contamination; small amounts (up to hundreds of ppm) may have come from additives in the original product. Typical contaminant levels found in used industrial lubricating oils are given in Table 3.

Table 2. Summary of Typical Vehicle Used Oil Contaminant Levels^a

Property or test, ppm	Used oil type ^b						
	GE	GT	DE	DT	HDE	HDT	RRE
arsenic	<1	<2.4	2	0.39	<1	0.38–1.59	<1
barium	1.0–43	11.6–32.6	1.5–6.4	9.7–76.4	1.5	<10	1.3–4.3
cadmium	0.5–3.4	1.0–5.0	0.7–3	0.27–1.9	0.8–4.5	0.51–1.48	12.0
chromium	0.8–23	2.67–5.0	1.8–7.1	2.45–7.0	1.5–8	0.89–2.43	1.1–43.3
lead	5.5–150	29–345	2.9–19.0	8.0–133	1–33	10.8	1.5–31.5
benzene	0.53–13.2	0.28–420	ND	19.3	NA	142	<2.5
trichloroethylene	<25	<50	ND	1.0	NA	NA	<2.5
perchloroethylene	<25	89–1700	ND	74	NA	NA	<2.5
trichloroethane	25	51–2100	ND	60	NA	NA	<2.5
tetrachloroethane	<25	<50	ND	<2	NA	NA	<2.5
benzo[b]fluoranthene	13–91	5–19	1.5	2.4–46	<5	NA	<5
benzo[k]fluoranthene	10–22	1.9–12	1.1	1.2	<5	NA	<5
benzo[a]pyrene	25–86	7.3–24	2.0	3.0	<5	NA	<5
PCBs	ND	ND	ND	ND	NA	NA	NA

^a Ref. 21.

^b GE = gasoline unleaded engine oil; GT = storage tank, automotive oils/fluids; DE = diesel engine oil; DT = storage tank, diesel maintenance; HDE = heavy-duty diesel engine oil; HDT = storage tank, heavy-duty diesel fleet; RRE = railroad diesel engine oil; ND = not detected; and NA = not analyzed.

Table 3. Summary of Typical Industrial Used Oil Contaminant Levels^a

Property or test, ppm	Used oil type ^b		
	HF	MWF	EIO
arsenic	3.26	2.0–21.5	<1
barium	1.4–460	0.3–8.1	<1
cadmium	1.4–10.1	1.3–4.8	<0.25
chromium	1.0–1.6	1.0–5.4	<1
lead	1.0–7.0	1.0–6033	1.0
benzene	ND	<5	<5
trichloroethylene	ND	<5	<5
perchloroethylene	ND	<5	<5
trichloroethane	ND	<5	<5
tetrachloroethane	ND	<5	<5
benzo[b]fluoranthene	<5	6	<5
benzo[k]fluoranthene	<5	<5	<6
benzo[a]pyrene	<5	<5	<5
PCBs	ND	ND	6.9

^a Ref. 21.
^b HF = hydraulic oil/fluids; MWF = metalworking oil/fluids; EIO = electrical insulating oils; and ND = not detected.

Technology

In the recycling of used oils to regain useful products, a number of stages are possible, depending on the original source of the used oil, the level of contamination, and the sophistication of the recycling technology utilized. Three levels of recycling are briefly described herein. In general, an oil cannot be recycled to a higher quality product than its original use, eg, a low or medium viscosity index (VI) industrial oil cannot be recycled into a high VI automotive engine oil. Further, it is important for cost-effective oil recycling to keep different oils segregated during the recycling process, because commingling of a high VI with a low VI used oil allows only recycling into low VI products.

Reprocessing. Combustion of used oil as burner fuel has often been condemned because it destroys a valuable resource and can cause substantial environmental pollution through widely dispersed distribution of metal oxides and stable organic contaminants. Further, these contaminants, sometimes referred to as products of incomplete combustion, may cause scaling of heat-transfer surfaces and fouling of burners and fuel-transfer lines if not handled or processed properly. However, under certain conditions and using suitable precautions, the recovery of the energy as heat is a valid used oil disposal option (2,5,12,23). The conditions and precautions have been described (7); typical regulatory standards can also be found (22).

Processing techniques for the recycling of used oil into fuel include pretreatment of the used oil to remove all or most of the contaminants that cause environmental or operational concerns. An alternative approach is to subject the used oil to minimal reprocessing followed by burning in specialized facilities using acceptable environmental control, ie, electrostatic precipitators, Venturi scrubbers, or a fabric filter baghouse (7). Lower level pretreatment options include settling or centrifugation, or both, to remove coarse solids and free water. Improved cleaning occurs when using heat and demulsifiers to remove water, volatiles, and most suspended solids. However, these methods do not significantly reduce the soluble contaminants or submicrometer-sized particulates which contribute to ash formation. Higher level treatments of used oil can be applied to further clean the oil, but are not normally employed for economic reasons.

Reclaiming. Used oils can be reclaimed within the user facility or sent outside the facility to a commercial reclaimer. The primary difference between these two options is usually the level of treatment available. Often the in-plant reclaiming facility is limited to gravity purification or settling, filtration, centrifuging, and heating to remove volatiles and water. A commercial reclaimer can usually perform all of these as well as providing clay treatment, chemical treatment such as acid or caustic, demulsification, distillation, and reformulation with additives, if necessary. A general description of such processes may include any or all of the following: (1) removal of solid particles by settling, centrifuging, or filtering; (2) neutralization of acidic components with clay or alkalis, and removal of resulting soaps by washing; (3) heating/distillation to remove volatile solvents, gasoline, and water; (4) clay contacting to remove oxygenated components and spent additives or for decolorization; (5) aerations and use of biocides to reduce bacterial levels; and (6) replenishment of additives.

The product of a reclaiming procedure is a reclaimed oil, which often meets original specifications for such uses as hydraulic fluids, gear lubricants, cutting and grinding oils, and metal-rolling lubricants. Perhaps the most important requirement for effective reclaiming is segregation of the used oils according to type. In general, a mixture of high quality and low quality oil can be reclaimed only as a low quality oil or as a fuel. Thus, the lack of segregation at the source may carry a high economic penalty.

Re-Refining. Petroleum refining processes are employed for re-refining used lubricating oil to produce clean, high quality lubricating base oil. These processes often include a pretreatment to reduce the impurity content by one or more of the following methods: application of heat, filtration, and treatment using acid, caustic, solvents, and/or other chemicals. In developed countries, pretreatment is usually followed by multistage vacuum distillation using clay or hydrogen finishing. Countries having less stringent environmental regulations continue to allow treatment using concentrated acid and large amounts of clay followed by plate-and-frame filtration. There is also scattered use of chemical pretreatment followed by hydrotreating, solvent extraction using clay or hydrogen finishing, and extensive treatment using only clay; the latter is limited to highly segregated used oil (24–27).

Many re-refining processes are described in the literature (23–28). The primary technology installed in developed countries involves some variation on multistage vacuum distillation followed by catalytic hydrotreating, as shown schematically in Figure 1 (2). Although other technologies have been proposed and even patented, the principal concern with many of these newer re-refining approaches is the lack of technical data and long-term experience establishing the quality of the re-refined base oil product. This subject is discussed in great detail in several publications (4,17,18,23–26,28). However, some economic assessments underscore the viability of the distillation–hydrotreat approach (27,29).

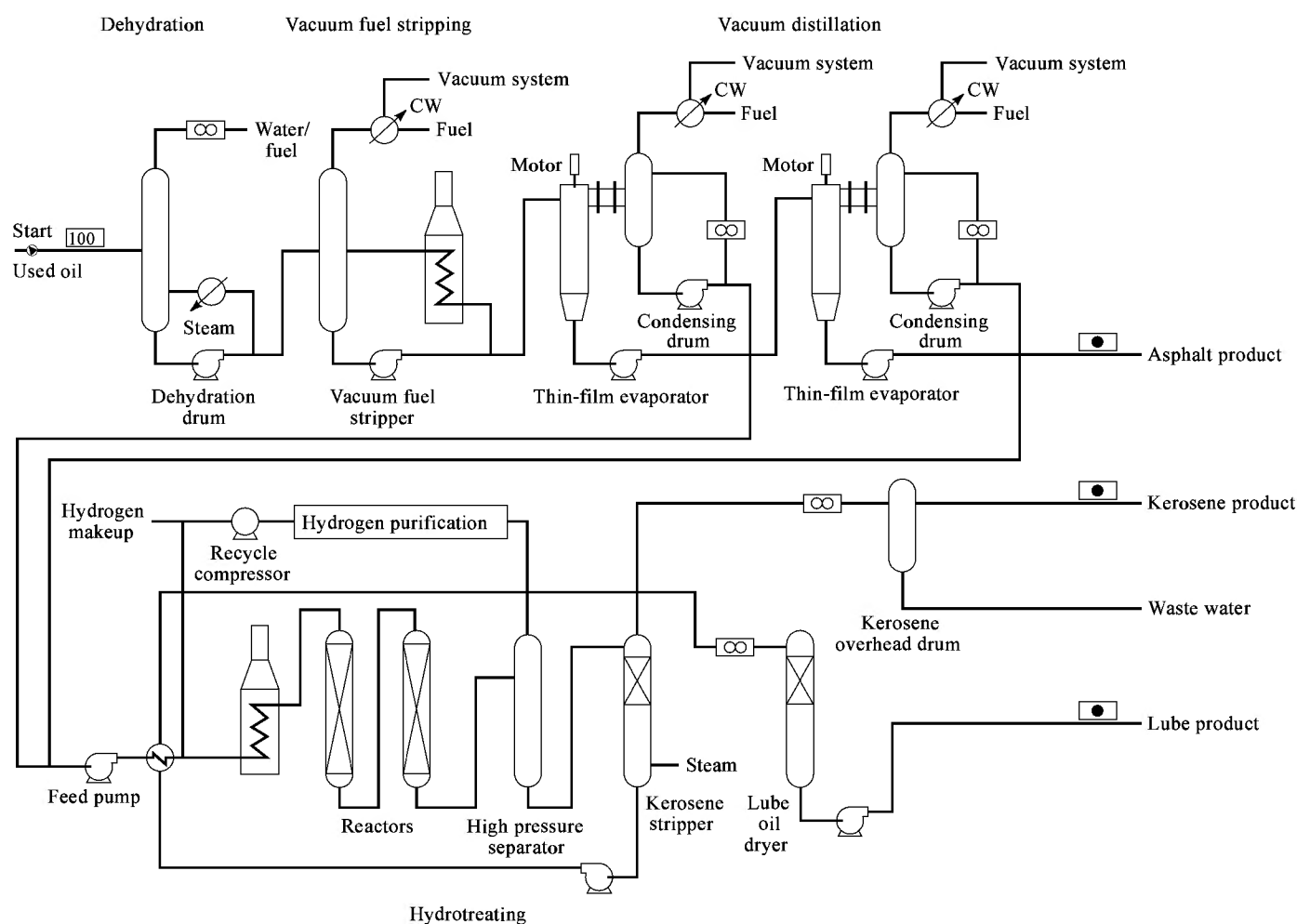


Fig. 1. Distillation-hydrotreat process, where CW = cooling water.

Regulations and Specifications

A significant source of concern for potential users of recycled petroleum products has been the lack of specifications or certifications related to the quality of the material and the consistency in producing oil of high quality. This perception of possible inferiority has been exacerbated by the reluctance of some equipment manufacturers to state whether they would honor warranties if recycled lubricants were used.

The literature is full of detailed evaluations of recycled petroleum products (2,6,17,18,23,28) and investigations into the environmental ramifications of the recycling processes themselves (30).

In addition, specifications have been produced by such organizations as the Canadian General Standards Board (31) and Green Seal (32); further activities are also occurring within the ASTM. The thrust of these actions is to require the recycled petroleum products to meet the same requirements as placed on their virgin equivalent, as well as to demonstrate that contaminants have been reduced to insignificant levels.

Based on the evidence that acceptable recycled petroleum products can be produced, there is a considerable legislative record encouraging the recycling of used oil. Starting with the Resource Conservation and Recovery Act in 1976 (20), used oil was held apart from the normal hazardous waste system because the oil was viewed as a valuable commodity. This was followed by the Used Oil Recycling Act in 1980 (33), which removed any federal requirement that lubricants containing re-refined base oil carry special labeling.

The U.S. EPA issued regulations implementing the legislation in stages, starting in 1988 with a Guideline for Federal Procurement of Lubricating Oils Containing Re-refined Oil (34) and concluding with Management Standards in 1992 (22). As part of the latter, the EPA decided not to list used oil as a hazardous waste in order to encourage its collection and recycling.

The Federal Trade Commission has announced a rule, effective November 30, 1995, that sets test procedures and labeling standards for recycled oil used as engine lubricating oil (35). The test procedures used are those contained in the Engine Oil Licensing and Certification System of the American Petroleum Institute (API) (36). The rule states in effect that if recycled oils meet the requirements of the API Certification System, such oils are substantially equivalent to new oil for use as engine oil. This federal rule preempts certain state recycled oil rules (35).

BIBLIOGRAPHY

"Recycling (Oil)" in *ECT* 3rd ed., Vol. 19, pp. 979-985, by D. A. Becker, National Bureau of Standards.

1. *Standard Terminology Relating to Petroleum, Petroleum Products, and Lubricants*, D4175, American Society for Testing and Materials, Philadelphia, Pa., 1988.
2. T. Pyziak and D. W. Brinkman, *J. Soc. Tribol. Lubr. Eng.*, 339 (May, 1993).
3. *Generation and Flow of Used Oil in the United States in 1988*, Temple, Barker, and Slone, Inc., 1989.
4. *Composition and Management of Used Oil Generated in the United States*, EPA/530-SW-013, Franklin Associates, Ltd., 1984.
5. Recon Systems, Inc., *Used Oil Burned as a Fuel*, Report No. SW-892, Environmental Protection Agency, Washington, D.C., 1980.
6. E. A. Grame and T. C. Bowen, Jr., *U.S. Army/Environmental Protection Agency Re-Refined Engine Oil Program*, AFLRL Report No. 98, U.S. Army MERADCOM, Ft. Belvoir, Va., 1978.
7. S. Chansky and co-workers, *Waste Automotive Lubricating Oil: Reuse as a Fuel*, Report No. EPA-600/5-74-032, Environmental Protection Agency, Washington, D.C., 1974.
8. *Fed. Reg.* 56, 48018 (Sept. 23, 1991).
9. J. W. Swain, Jr., *Recycled Oils for Fuel Uses*, unpublished report to the National Bureau of Standards, Washington, D.C., Jan. 1978.
10. Report to the Congress, *Waste Oil Study*, Environmental Protection Agency, Washington, D.C., Apr. 1974.
11. *Environ. Sci. Technol.* 6, 25 (Jan. 1972).

12. E. E. Berry, L. P. MacDonald, and D. J. Skinner, *Experimental Burning of Waste Oil as a Fuel in Cement Manufacture*, Environment/Canada Report No. EPS 4-WP-75-1, Ottawa, Canada, June 1975.

13. *Energy from Used Lubricating Oils*, API Pub. No. 1588, American Petroleum Institute, Washington, D.C., Oct. 1975.

14. P. Cukor, *Economic and Environmental Assessment of Used Oil Recovery and Disposal for Ontario*, report by Teknekron under contract to the Ontario Energy Management Program, Ontario, Canada, Mar. 1976.

15. N. Suprenant, J. Sahagian, and R. Hall, *Waste Oil Recovery and Refuse Program: Residue Management*, GCA/Technological Report No. GCA-TR-75-3-G, prepared for Maryland Environmental Services, by GCA/Tech., Bedford, Mass., Apr. 1975.

16. M. L. Whismann, J. W. Goetzinger, and F. O. Cotton, *Waste Lubricating Oil Research*, Bartlesville Energy Research Center, Bartlesville, Okla., 1974–1977.

17. D. A. Becker, ed., *Measurements and Standards for Recycled Oil-I*, NBS Special Publication 488, National Bureau of Standards, Washington, D.C., Aug. 1977.

18. D. A. Becker, ed., *Measurements and Standards for Recycled Oil-II*, NBS Special Publication 556, National Bureau of Standards, Washington, D.C., Sept. 1977.

19. Toxic Substance Control Act, Public Law 94-469, U.S. Congress, Washington, D.C., Oct. 11, 1976.

20. Resource Conservation and Recovery Act, Public Law 94-580, U.S. Congress, Washington, D.C., Oct. 21, 1976.

21. *Fed. Reg.* **56**(184), 48000 (Sept. 23, 1991).

22. *Fed. Reg.* **40**(279).

23. L. Y. Hess, *Reprocessing and Disposal of Waste Petroleum Oils*, Noyes Data Corp., Park Ridge, N.J., 1979.

24. M. L. Whisman, *Lubr. Eng.* **35**, 249 (1979).

25. D. W. Brinkman, *Lubr. Eng.* **43**, 324 (1987).

26. D. J. McKeagan, *Lubr. Eng.* **48**, 418 (1992).

27. D. W. Brinkman, *Oil Gas J.*, 60 (Aug. 19, 1991).

28. D. A. Becker, ed., *Joint Conference on Measurements and Standards for Recycled Oil/ Systems Performance and Durability*, NBS Special Publication 584, National Bureau of Standards, Washington, D.C., Nov. 1980.

29. R. Arner, *Used Oil Recycling Markets and Best Management Practices in the U.S.*, North Virginia Planning District Commission, Annandale, Va., Oct. 1992.

30. N. F. Surprenant and co-workers, *The Fate of Hazardous and Nonhazardous Wastes in Used Oil Disposal and Recycling*, DOE/BC/10375-6, Department of Energy, Washington, D.C., 1983.

31. *Re-Refined Base Oils*, Canadian General Standards Board, Canada.

32. *Environmental Standard for Re-Refined Engine Oil*, Standard GS-3, Green Seal, Inc., Washington, D.C., 1992.

33. Used Oil Recycling Act of 1980, Public Law 94-463, U.S. Congress, Washington, D.C.

34. *Guideline for Federal Procurement of Lubricating Oils Containing Re-Refined Oil*, 42USC252, U.S. Congress, Washington, D.C., June 30, 1988.

35. *Fed. Reg.* **16**(311), 55414–55422 (Oct. 31, 1995).

36. *Engine Oil Licensing and Certification System*, 13th ed., API Publication 1509, American Petroleum Institute, Washington, D.C., 1995.

General Reference

Energy Policy and Conservation Act, Public Law 94-163, U.S. Congress, Washington, D.C., Dec. 22, 1975.

Donald A. Becker
National Institute of Standards and Technology

Dennis W. Brinkman
Safety-Kleen Corporation

PAPER

Paper recycling mills encompass a range of unit operations. The choice and sequence of these operations is determined by the types of recovered paper being processed, the types of paper products produced, availability of process water, economic considerations, and environmental considerations. There are a wide variety of recovered paper types (Table 1) and grades within each type (1). The types of recovered paper being processed determines the contaminants that must be removed. The degree to which these contaminants are removed is determined by the types of paper products being produced. Contaminants include ink, adhesives and glue, rosin, wax, starches and gums, coatings, paper fillers, styrofoam, and plastic from bags and tape. Adhesives, glue, and waxes are grouped together and called stickies (2,3).

Table 1. Recovered Paper Sources

Source	Abbreviation	Recycled paper products
Post-consumer		
old corrugated containers	OCC	linerboard, boxboard, corrugating medium, tissue, paper towels
linerboard (cardboard boxes)	DLK	linerboard, boxboard, corrugating medium
old newspapers	ONP	newsprint, boxboard, tissue, paper towels
old magazines	OMG	magazine newsprint, tissue, towels
computer printouts	CPO	printing and writing paper, tissue, paper towels
ledger		printing and writing paper, boxboard, tissue, towels
mixed office paper	OWP or MOW	printing and writing paper
Pre-consumer ^a		
boxboard cuttings		linerboard, boxboard
envelope cuttings		envelopes, tissue, paper towels
printer trims		depends on original paper type

^a Commercial–industrial–converter material.

Inefficient removal of ink can cause low recycled paper brightness and the appearance of visible specks on recycled paper sheets. Stickies can also be visible on the sheet and can cause additional problems. Stickies can interfere with efficient paper machine operation by sticking to wires, felts, paper machine head boxes, and dryer cans. This plugging can reduce paper machine drainage rates forcing mill operators to run machines at lower speeds thus reducing paper production. Periodically shutting down the mill to clean felts also reduces production. Some stickies can remain in the paper causing it to adhere to itself when taken up on a paper roll. This can cause press room breaks interfering with printing operations. Thus, efficient stickies removal is essential.

Important unit operations in these mills include pulping, high density cleaning, screening, forward cleaning, reverse cleaning, washing, flotation, dispersion and kneading, bleaching, water treatment, and sludge handling. These operations are designed to remove contaminants from the pulp. The effectiveness of these separation operations is determined by contaminant particle size, geometry, density, and surface chemistry. This is illustrated in Figure 1 for ink particles (4). The unit operations most effective for various types of other contaminants (2) are summarized in Table 2. Correct choice of process chemicals can improve the effectiveness of some of the unit operations listed in Figure 1 and Table 2.

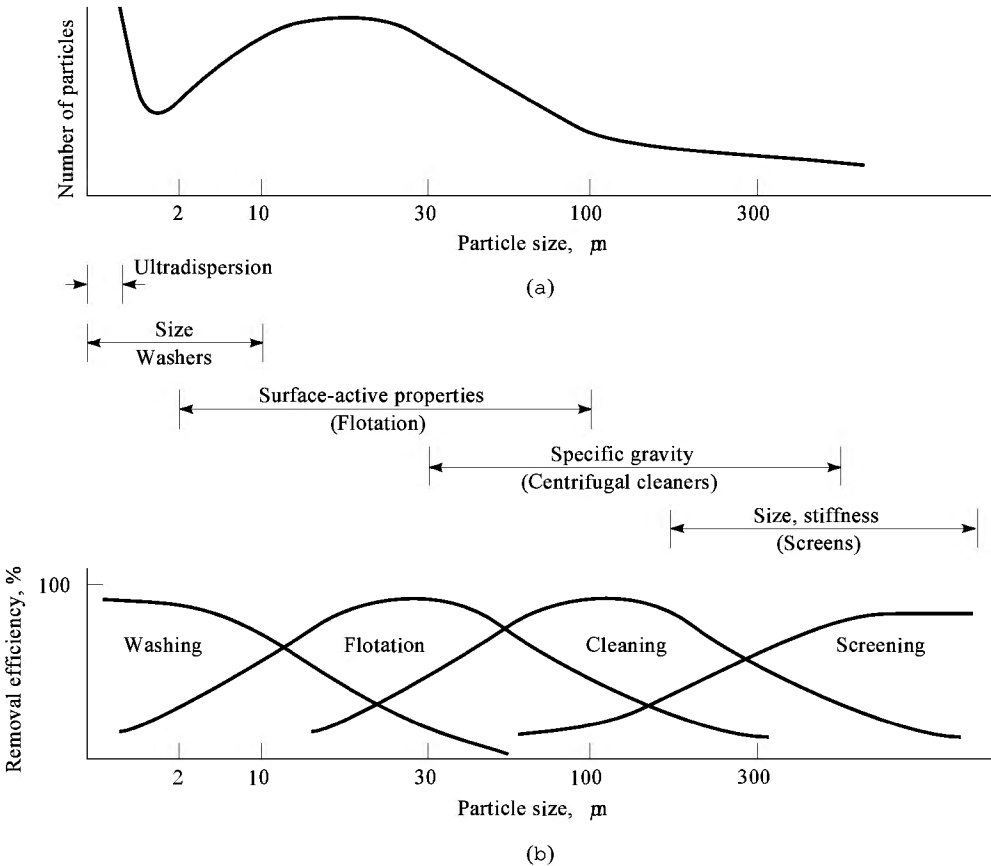


Fig. 1. Ink removal effectiveness of unit operations as a function of ink particle size: (a) particle size distribution in pulper; (b) unit removal efficiency (4).

Table 2. Unit Operations for Removal of Various Contaminants from Recovered Paper^a

Contaminant	Particle size	Specific gravity	Unit operation
metal bolts, screws, wire, staples, paper clips, sand, dirt, glass	very large	usually >1.0	perforated plate in pulper covering exit flow line, coarse screens
wood chips, string, nylon mesh, rags, kraft shipping and wrapping paper, laminated paper, sealing tape, cellophane from envelope windows, poly-ethylene-coated board, mis-cellaneous plastics from styrofoam cups, packaging, etc	medium large	usually <1.0	coarse screens, fine screens
wood shives, flakes from paper coatings ^b	medium small	usually <1.0	fine screens, forward cleaners
paper fillers ^c , mineral waxes, some hot-melt adhesives, high viscosity adhesives ^d	small	>1.0	forward cleaners, washing, flotation, mechanical dis-persion
low and medium viscosity adhesives ^e	small	<1.0	reverse cleaners, washing, flotation

^a Ref. 2.
^b 1.5–5.0 mm² area.
^c For example, clays or titanium dioxide.
^d For example, asphalt.
^e For example, hot melts, waxes, and natural or synthetic resins.

Representative sequences of unit operations are presented in Figure 2 for a wash deinking mill processing old newspapers to make newsprint, a modern mill processing mixed office papers to produce printing and writing paper, a flotation–wash mill processing a mixture of old newspapers and old magazines to make newsprint, a mill processing old paper to make tissue products, and a mill recycling old corrugated containers.

Newsprint wash deinking mill	Office paper deinking mill	Newsprint/magazine deinking mill	Tissue mill	Old corrugated container mill
Pulper	Pulper	Pulper	Pulper	Pulper
HD cleaner	HD cleaner	HD cleaner	HD cleaner	HD cleaner
Coarse screens	Coarse screens	Coarse screens	Coarse screens	Coarse screens
Fine screens	Fine screens	Flotation cells	Optional	Dilution
Forward cleaners	Thickener	Forward cleaners	flotation cells	Forward cleaners
Reverse cleaners	Kneader	Reverse cleaners	Reverse cleaners	Reverse cleaners
Washer	Flotation cells	Washer	Forward cleaners	Fractionation
Bleach tower	Reverse cleaners	Thickener	Washer	Thickener
	Forward cleaners	Dispersion unit	Bleaching	Dispersion unit
	Washer	Flotation cells		
	Bleach tower	Reverse cleaners		
		Washer		
		Bleach tower		

Fig. 2. Simplified process designs of paper recycling mills. HD = high density.

Unit Operations

Pulping. The first of the unit operations is pulping. Pulping disintegrates paper into individual fibers dispersed in water. It may be a batch or continuous process. Added chemicals promote ink detachment from fibers and adjust the dispersed ink particle size. Newsprint deinking mills commonly use continuous pulping. Most mills producing tissue products or printing and writing paper from deinked pulp use batch pulping.

Low consistency pulping (3–6% solids) is common in newsprint and many tissue mills. Medium (6–12%) and high consistency pulping (12–18% solids) is common in mills deinking office papers. Pulping temperature is typically 40–55°C, the pH is usually 9.0–10.5, and process time ranges from 4 to 60 minutes.

Both mechanical and chemical action promote ink detachment from cellulose fibers during pulping. Mechanical action includes interfiber abrasion and fiber flexing and bending. Chemical action includes fiber swelling and surfactant-promoted ink particle emulsification and solubilization.

Fiber swelling is promoted by high pH. For this purpose, sodium hydroxide is often added to the pulper. As much as 3% based on dry paper weight may be used (5). However, more typical dosages are 0.8–1.5%. Newsprint and magazines are commonly pulped at pH 8–10. In office paper deinking, pulper pH is sometimes as high as 10–11 but may be as low as 7–8. High pH can promote yellowing of lignin-containing pulps, usually groundwood or thermomechanical pulp. To minimize this, a bleaching agent, usually hydrogen peroxide, is often added to the pulper. Hydrogen peroxide treatment levels up to 2% of the dry paper weight are used (5). Chelants are added to retard hydrogen peroxide decomposition promoted by multivalent metal ions present in the process water. The most commonly used are sodium silicate, ethylenediaminetetraacetic acid (EDTA), and diethylenepentaaminetetraacetic acid (DPTA). EDTA and DTPA dosage levels are about 0.15–0.4% relative to dry paper weight. Sodium silicate, which also may be used to control pH, is used at dosages of 1.0–3.0% based on dry paper weight (5).

Surfactants (qv) are added to the pulper to promote ink particle detachment from fibers and dispersion of detached ink in the process water. Dosages are 0.10–1.5% relative to dry paper weight (5). Mechanisms of ink removal are similar to those proposed for liquid soil removal from fabrics (6). Toner inks are thought to behave similarly to solid soils (6,7). Some researchers have proposed using enzymes to promote deinking (8,9). The enzyme is thought to function by detaching ink-containing fibrils from cellulose fibers during pulping.

Surfactants also can provide some control over the particle size of detached ink. The effectiveness of different deinking unit operations in separating dispersed ink from cellulose fibers varies with ink particle size (see Fig. 1). In deinking office paper containing toner ink-printed paper, some deinking agents promote aggregation of dispersed ink into three-dimensional ink particles (10–13). These are more easily removed by fine screens and some types of cleaners than the two-dimensional toner ink flakes usually formed during pulping. Thus, deinking process engineering can determine the type of surfactant used in the pulper.

Ink particle redeposition on cellulose fibers can reduce deinked paper brightness. Should larger ink particles redeposit on fibers, visible ink specks may result. Sodium silicate is often added to the pulper to act as a dispersant and reduce this redeposition (6,14). Up to 5% based on dry paper weight may

be used (5). The problem appears most severe with flexographic newsprint ink because of the very small particle size of the dispersed ink (14). These very small ink particles are not difficult to detach from cellulose. However, they redeposit on fibers readily (15). Use of laundering antiredeposition agents such as carboxymethylcellulose and sodium poly(acrylate) can increase deinked sheet brightness (16). Presumably, these chemicals function by reducing dispersed flexographic ink particle redeposition onto cellulose. Removal of ink particles from process water during water clarification for recycle is also important in reducing ink redeposition.

Mild pulping conditions preserve stickies as relatively large particles permitting their later removal by screens and mechanical cleaners. For example, wax is a common contaminant in mills recycling old corrugated containers. Pulping at temperatures less than 50°C (120°F) keeps the wax from melting, making it easier to remove using fine screens and mechanical cleaners.

Another technique to reduce the problems caused by stickies is to use additives to reduce the tackiness of these particles. This prevents their later reagglomeration and attachment to paper machine surfaces. These additives are usually added to the pulper. The most common is talc (17) usually added to the pulper in repulpable bags. Emulsified talc is also sometimes added to the pulp just before the pulp encounters high shear. Organic polymers (18) such as a polyvinylpyrrolidinone (PVP) copolymer (19) have also been reported to reduce the tackiness of stickies.

Dispersants (qv) have been added to the pulper to maintain stickies in a colloidal state. The small particle size reduces the problems stickies cause on the paper machine and in paper products. Among the chemicals that have been used are fatty alcohol ethoxylates, alkylphenol ethoxylates, lignosulfonates, and naphthalene sulfonates (18).

High Density Cleaning and Screening. Usually as the first step after pulping (20), cleaners remove high and medium density large particles, eg, rocks and dirt, nuts, bolts, nails, paper clips, and other objects often found in wastepaper. Centrifugal forces separate the less dense cellulose fibers from these heavy objects. Large objects also are removed using a perforated plate called a pulper detritator.

Screening usually follows high density cleaning (21,22). Screening removes relatively large ink (see Fig. 1) and usually low density contaminants from the pulp slurry. Coarse screens are fitted with holes which permit the passage of cellulose fibers and liquids while holding back large particles. Hole size ranges from 6 to 20 mm (21). Fine screens are fitted with slits as small as 0.15–0.30 mm in width. These separate smaller ($>250\ \mu\text{m}$) contaminant particles and toner ink particles from the pulp. These contaminants include unpulped paper, plastic, and large adhesive particles from envelopes and labels. Fine screening also can remove some large toner ink particles. Careful design of the rotor in these screens can minimize the buildup of a mat of fiber reducing operating rates and separation efficiency.

Washing. Washing removes dispersed ink particles from the pulp slurry (23). This process is comparable to home laundering in many ways. The surfactant is added to the pulper at a dosage level of 0.25–1.5% relative to dry paper weight. Strong dispersants such as alcohol ethoxylates and alkylphenol ethoxylates perform well in wash deinking. These surfactants promote ink particle dispersion by increasing the hydrophilic nature of the ink particles on which they absorb. Washing is most effective in removing small, dispersed ink particles such as letterpress, offset, and flexographic newsprint inks. It is less effective on large, poorly dispersed ink particles such as toner inks from photocopiers and laser printers and ultraviolet- and heat-set inks.

The optimum surfactant hydrophilic:lipophilic balance (HLB) for wash deinking is dependent on ink composition. Surfactants with a HLB of about 14.5 provide the highest deinked newsprint brightness (24). The optimum deinking surfactant HLB for ledger inks is 13–14, whereas that for toner inks is 10–11 (25).

Water removed from the pulp slurry during washing passes through a mat of paper fibers. As more water is removed from the pulp, the pulp consistency at the washer discharge increases. Commercial washers can be classified on the basis of their discharge pulp consistency (23). Low consistency ($<8\%$) washers employ sidehill screens and gravity deckers. Intermediate consistency (8–15%) washers are high speed belt washers, inclined screw extractors, and vacuum filters. High consistency washers ($>15\%$) are screw presses and belt presses. Cellulose fiber loss is a function of washer design and pulp discharge consistency (23).

Typical papers processed using wash deinking are 100% old newspaper and sorted office paper from which toner ink-printed paper has been removed. The effluent from washers is heavily laden with ink, mineral coating and filler particles, and small cellulose fibers. As a result, it can be difficult to clarify.

Flotation. Flotation (qv) is used alone or in combination with washing and cleaning to deink office paper and mixtures of old newsprint and old magazines (26). An effective flotation process must fulfill four functions. (1) The process must efficiently entrain air. Air bubble diameter is about 1000 μm . Typically air bubbles occupy 25–60% of the flotation cell volume. Increasing the air:liquid ratio in the flotation cell is said to improve ink removal efficiency (27). (2) Ink must attach to air bubbles. This is primarily a function of surfactant chemistry. Air bubbles must have sufficient residence time in the cell for ink attachment to occur. (3) There must be minimal trapping of cellulose fibers in the froth layer. This depends on both cell design and surfactant chemistry. (4) The froth layer must be separated from the pulp slurry before too many air bubbles collapse and return ink particles to the pulp slurry.

Pulp slurry consistency during flotation is typically 0.7–1.2%. Air enters the pulp slurry as bubbles. Surfactants adsorbed on ink particles render them hydrophobic promoting adsorption onto air bubbles. Thus, optimum HLB values are significantly lower for flotation deinking surfactants than for wash deinking surfactants. Flotation deinking surfactants are added either to the pulper or immediately before flotation. Adsorbed ink and mineral coating and filler particles rise with the bubbles to the top of the flotation cell. These solids plus small cellulose particles are trapped in the foam layer. The surfactant stabilizes the foam long enough for it to be removed before many bubbles collapse and return ink to the pulp slurry.

Early flotation deinking surfactants were fatty acids, often referred to as collectors. They require calcium ions to function and a minimum water hardness of about 90 ppm (as calcium carbonate) is needed (1). The calcium salt can be generated *in situ*, usually by addition of a soluble Ca^{2+} salt such as calcium chloride. Process water naturally containing high levels of hardness ions may not require addition of a calcium salt. Calcium carbonate filler particles in magazine paper can also serve as a calcium source. Fatty acids are added just before flotation at dosage levels up to 1.0% based on dry paper weight. This is two to five times that of more recently developed flotation deinking surfactants, ie, proprietary alkoxyates of fatty alcohols or fatty acids containing both ethoxy (EO) units and propoxy (PO) units (28,29). Patents suggest the optimum products are synthesized by adding an approximately 2:1 by weight mixture of EO and PO to an alcohol or carboxylic acid substrate such as stearic acid (30,31). However, the situation is complex as EO:PO ratios ranging from 1:2 to 4:1 have been claimed (32,33). Alkoxyates in which the EO and PO are added separately to form two or more blocks can also be used (34). Deinking efficiency of twin-tailed hydrophobe alkoxyates is said to be superior to single-chain hydrophobe alkoxyate surfactants (28). Ethylene oxide–propylene oxide copolymers are also used in flotation deinking operations (35).

Sodium silicate wetting, emulsifying, penetrating, and other surface-active properties do not appear to impact ink removal efficiency in flotation (36). Flotation can also remove some of the paper filler and coating particles dispersed in the pulp. Addition of certain cationic organic polymers such as poly(diallyldimethylammonium chloride) to pulp improves the removal efficiency of these particles during flotation (37,38).

Mechanical Cleaning. A cleaner is a hydrocyclone device utilizing fluid pressure to create rotational fluid motion (20). Pulp is introduced tangentially near the top of the cleaner. Contaminants denser than water such as chemically treated toner inks and sand migrate toward the outer wall of the cleaner and exit in a separate (reject) stream. For most forward cleaners, optimal ink removal efficiency is obtained at a pulp consistency of 0.2–0.3%. Most forward cleaners' deinking efficiency declines at pulp feed consistencies greater than 0.4%. However, a cleaner said to be efficient at 1.2% pulp consistency has been reported (39).

Reverse cleaners operate on the same principles as forward cleaners (20). Contaminants less dense than water migrate toward the center of the cleaner and exit as a separate (reject) stream from the pulp slurry. Reverse cleaners are used to remove adhesive and plastic particles as well as paper filler particles and lightweight particles formed from paper coatings.

Cleaners are most efficient on relatively large particles, 80–300 μm in diameter (see Fig. 1). Flat toner ink particles can fragment during processing, therefore it is probably best to locate mechanical cleaners early in the sequence of office paper deinking unit operations (40).

Forward and reverse cleaning are considered to be purely mechanical processes. However, use of certain deinking agents can facilitate removal of toner inks using forward cleaners by increasing ink particle size and or density in the pulper (13). Proprietary surfactants have been found to promote the formation of three-dimensional dispersed toner particles (10). These have a higher apparent density than the flat ink flakes usually formed during pulping. The three-dimensional character thus facilitates ink removal in forward cleaners (10,41).

Dispersion and Kneading. These are mechanical processes designed to reduce dispersed ink particle size through fiber–ink abrasion processes (42,43). Fiber–fiber abrasion can detach additional ink from fibers. Dispersion is often performed at elevated temperatures above the softening point of toner inks (ca 65°C), uv-cured inks (85–120°C), and adhesive particles (85–110°C). Particle softening can aid in reducing particle size. Kneading may be performed at ambient temperature. Both processes are performed at high consistency, up to 30° o by weight cellulose fiber (44), and reduce the number of visible ink particles.

Dispersion and kneading are often used for pulp containing toner inks because these inks tend to form large particles during pulping. Dispersion and kneading reduce toner ink particle size below the visible range. With pulp made from old newsprint and magazines, dispersion is sometimes used after flotation and washing to improve optical homogeneity of paper made from the deinked pulp. Washing and or flotation are often performed after dispersion or kneading to remove the small ink particles formed in these processes. This removal significantly increases deinked pulp brightness (44). Flotation can also further reduce the number of visible ink particles.

Dispersion at temperatures of 90–110°C is a common final step in European mills processing wax-coated old corrugated containers. Dispersion temperatures less than 90°C are reported to reduce wax particle size to improve pulp drainage properties on paper machines while improving paper strength (45). Dispersion has been used to reduce hot-melt adhesive, plastic coating, and asphalt particle size. These low density particles can then be removed from the pulp by flotation (46).

Bleaching. Bleaching was first developed for cellulose fibers prepared directly from wood (47) (see BLEACHING AGENTS, PULP AND PAPER). In paper recycling, bleaching agents are used both during pulping and in separate bleaching operations performed after removal of inks and other contaminants from the pulp (48). Bleaching during pulping to retard fiber darkening at high pH is discussed in the section on pulping.

Bleaching as a separate operation whitens fibers and removes the coloring effect of dyes. As practiced in paper mills, bleaching is a multistage operation using different bleaching agents in each stage.

There are two types of bleaching: oxidative and reductive. Oxidative bleaching cleaves carbon–carbon double bonds thus destroying chromophores. Reductive bleaching reduces carbon–carbon double bonds to single bonds. The loss of extended conjugation in the chromophore leads to loss of color.

Common oxidative bleaches are hydrogen peroxide, sodium hypochlorite, chlorine dioxide, oxygen, and ozone. Although it is a very cost-effective bleaching agent, sodium hypochlorite is not used extensively to whiten deinked pulp due to environmental concerns. Reductive bleaches include sodium hydrosulfite and formamidine sulfinic acid (FAS). Improved cost effectiveness is claimed when FAS is generated in the pulp slurry by reaction of hydrogen peroxide and thiourea (49). Oxygen gas in combination with an alkaline agent such as sodium hydroxide has been reported to reduce the tackiness of stickies (50).

Refining and Fractionation. These processes are used to alter and select cellulose properties so the final sheet has the desired properties (51). Properties of recycled fibers differ from those of fibers prepared directly from wood. For example, recovered chemical fibers have lower freeness, an apparent viscosity leading to different water drainage characteristics on paper machines. Recovered fibers also have increased apparent density, lower sheet strength, increased sheet opacity, inferior fiber–fiber bonding properties, lower fiber swelling, lower fiber flexibility, lower water retention, reduced fiber fibrillation, and much lower internal fiber delamination.

Refining is used to develop the desired pulp drainage properties and control the following sheet properties: bulk and density, strength, surface smoothness, porosity, and printing characteristics.

Fractionation separates fines and short, weak cellulose fibers from longer, stronger cellulose fibers (52). Its primary application is in processing old corrugated containers into new packaging products.

Water Clarification. Process water that needs to be clarified comes from several different sources in the recycling mill: rejects from screens and mechanical cleaners; rejects from washers, thickeners, and flotation cells; water that drains from the pulp as it is converted into paper on the paper machine (white water); and water from felt washers. These waters contain different dissolved chemicals and suspended solids and are usually processed separately.

Mills have three principal options in handling their process water: (1) discharge the water to a municipal treatment facility, (2) use sedimentation tanks to allow solids to settle out of the liquid, and (3) dissolved air flotation. Dissolved air flotation is faster than settling and the water loses less heat before being returned to mill operations. The sludge removed from the dissolved air flotation unit is higher in consistency than sludge from settling tanks.

Water from screens, cleaners, washers, thickeners, and flotation cells contain relatively high levels of ink. These waters also contain valuable chemicals, ie, sodium hydroxide and surfactants. Recycle of this water can save up to 10° o in chemical costs.

Customarily, combinations of cationic and anionic flocculants are used in deinking process water clarification. Typically, a low molecular weight cationic polymer is first added to neutralize negative charge on suspended solids. For economic reasons, the polymer is typically an ammonium salt. Then a high molecular weight anionic polymer is added to flocculate the suspended solids. This is usually a copolymer of acrylamide or a chemically modified polyacrylamide. The nature of the deinking surfactant can have a significant effect on the efficiency of flocculants in clarifying deinking process water (13).

Inorganic chemicals may also be used. Bentonite may be used as a flocculant in combination with polymer treatment. Alum, once a common coagulant, is less used because its concentration can build up in recycle water. Alum often binds ink to fibers and increases the difficulty of deinking. Removal of the very small flexographic ink particles in process water is difficult. Ultrafiltration (qv) has been proposed for removing these very small dispersed ink particles (53).

Rejects and Sludge Handling. Sludge from water clarification contains water, inks and solid pigments, dispersed adhesive particles, small plastic particles or wax, short cellulose fibers, paper filler and coating particles, and large solid materials, eg, rocks, dirt, wire, ceramics, etc.

Efficient dewatering minimizes sludge volumes sent to landfills thus reducing hauling costs and tipping fees. Efficient sludge dewatering also increases the efficiency of heat utilization during incineration. Inclined screw thickeners are often used to thicken the rejects from dissolved air flotation clarifiers. Reciprocating piston presses are also used. Polymers are often used to improve drainage across a sludge press thus increasing the consistency of the final sludge. Sludge applications under investigation include use in bricks and as a fertilizer spray on fields and roadsides.

Economic Aspects

In 1993, nearly 36 million tons of paper were recovered in the United States, twice as much as in 1980 (54). For the first time, more paper was recovered in the United States than landfilled. As a result, 11 million fewer tons of paper were landfilled in the United States in 1993 than in 1987. This saved more than 69 × 10⁶ m³ (90 × 10⁶ yd³) of landfill space. In 1995, recovered paper accounted for 31.5° o of the fiber used to manufacture 84.1 million metric tons of paper products (54). Annual capital spending for paper recycling projects from 1995 to the year 2000 is estimated to average \$2 billion (55). The American Forest & Paper Association (AF&PA) estimates U.S. consumption of recovered paper will increase 4.9° o per year through the year 2000, nearly twice the total paper industry capacity growth rate (56). Consumption of recovered paper in U.S. mills in 1997 is estimated at 35.6 million tons (57).

Paper recycling has greatly increased in Canada, from one mill making recycled-content newsprint in 1989 to 60 paper recycling operations in 1995 (58). Canadian consumption of recovered paper is about 4 million t yr, much of it imported from the United States. Paper recycling continues to grow worldwide, particularly in Europe and the Pacific Rim. Worldwide use of recycled paper is expected to increase from nearly 75 million tons in 1988 to 130 million tons in 2001 (58).

This growth will substantially increase demand for surfactants, bleaching agents, complexing agents, and other chemicals used in paper recycling (59). Deinking chemical demand is forecast to increase over 9° o a year from 55 million t in 1990 to 134 million t in the year 2000 (60). These chemicals include surfactants, fatty acids, silicates, sodium hydroxide, and others (61). Increases in paper recycling and changes in bleaching technology are cited as the reasons EDTA chelating agent use is expected to increase 3–5° o annually (61). Defoamer use is also increasing due to the growth in paper recycling (61). Deinking surfactant usage in Japan was 9,000–10,000 metric tons per year in 1992 and continues to increase (62).

Recycled paper contains more fines, short fibers, and anionic trash. This will increase demand for process chemicals such as drainage aids and both wet- and dry-strength additives (43).

BIBLIOGRAPHY

"Paper" under "Recycling" in *ECT* 3rd ed., Vol. 21, pp. 986–992, by J. H. Robins and J. R. Grant, Mytec, Inc.

1. 1995 *PIMA Catalog*, Paper Industry Management Association, Mt. Prospect, Ill., pp. 3E, 4E, R14–18.
2. W. F. Carr, in *Recycling Paper from Fiber to Finished Product*, Vols. 1 and 2, TAPPI Press, Atlanta, Ga., 1990.
3. M. Doshi, *Prog. Paper Recyc.* **4**(2), 103 (1995).
4. M. A. McCool and L. Silveri, *Tappi J.* **70**(11), 76 (1987).
5. A. Renders, *Proceedings of the TAPPI Pulping Conference*, TAPPI Press, Atlanta, Ga., 1992, p. 233.
6. J. K. Borchardt, *Coll. Surf. A.; Physiochem. Eng. Asp.* **88**, 13 (1994).
7. J. B. W. Morsink and G. J. R. Daane, in Ref. 5, p. 963.
8. H-J. Putz, K. Renner, L. Gttsching, and O. Jokinen, *Proceedings of the TAPPI Pulping Conference*, TAPPI Press, Atlanta, Ga., 1994, p. 877.
9. T. Welt and R. J. Dimas, *Prog. Paper Recyc.* **4**(2), 36 (Feb. 1995).
10. J. K. Borchardt, D. W. Matalamaki, V. G. Lott, and G. A. York, *Prog. Paper Recyc.* **4**(1), 44 (Nov. 1994).
11. T. Rhodes and L. Ferguson, *Proceedings of the TAPPI Recycling Symposium*, TAPPI Press, Atlanta, Ga., 1993, p. 123.
12. C. R. Olson, J. D. Hall, and I. J. Phillippe, *Prog. Paper Recyc.* **2**(2), 49 (1993).
13. W. D. Darlington, *Tappi J.* **72**(1), 35–38 (1980).
14. T. Ali, F. McLellan, J. Adwinata, M. May, and T. Evans, *J. Pulp Pap. Sci.* (1), J3 (Jan. 1994).
15. L. Gttsching and H-J. Putz, *Proceedings of the TAPPI Recycling Symposium*, TAPPI Press, Atlanta, Ga., 1994, p. 207.
16. J. K. Borchardt, K. H. Raney, P. G. Shpakoff, D. W. Matalamaki, and D. R. Denley, in Ref. 8, p. 1067.
17. **U.S. Pat. 5,336,372** (Aug. 23, 1994), L. D. Markham and N. R. Srivatsa (to International Paper Co.).
18. J. Ward, D. Hensel, and J. Patterson, in Ref. 15, p. 67.
19. **Can. Pat. 2,091,273** (Oct. 29, 1993), T. Ling (to Betz Laboratories, Inc.).
20. K. Merriman, in R. J. Spangdenberg, ed., *Secondary Fiber Recycling*, TAPPI Press, Altanta, Ga., 1993, p. 101.
21. T. Bliss, in Ref. 20, p. 125.
22. C. M. Vitori, *Pulp Pap. Can.* **94**(12) 109 (1993).
23. R. G. Horacek and W. Forester, in Ref. 20, p. 163.
24. L. L. Turai and L. D. Williams, *Tappi J.* **60**(11), 167 (Nov. 1977).
25. J. K. Borchardt and J. H. Rask, *Proceedings of the TAPPI Pulping Conference*, TAPPI Press, Atlanta, Ga., 1993, p. 839.
26. M. A. McCool, in Ref. 20, p. 141.
27. M. W. Gilkey, P. Seifert, and T. O. Conte, in Ref. 15, p. 163.
28. K. Masamizu, Y. Tai, M. Hagiwara, and T. Ukigai, in Ref. 15, p. 39.
29. N. N-C. Hsu and I. J. Vazquez, in Ref. 15, p. 79.
30. Jpn. Kokai Tokkyo Koho 06 49,790 (Feb. 24, 1994), Y. Ishibashi, Y. Myauchi, M. Inoe, and T. Edo (to Kao Corp.).
31. Jpn. Kokai Tokkyo Koho 05,222,686 (Aug. 31, 1993), T. Shiroishi and co-workers (to Kao Corp.).
32. **Ger. Pat. 4,217,907** (Dec. 3, 1992), Y. Ishibashi and H. Urushibata (to Kao Corp.).
33. **Jpn. Kokai Tokkyo Koho 04,289,287** (Oct. 14, 1992), Y. Ishibashi and H. Urushibata (to Kao Corp.).
34. **Jpn. Kokai Tokkyo Koho 05,263,379** (Oct. 12, 1993), K. Kato (to Nitsushin Kagaku Kenkyusho KK).
35. **Can. Pat. Appl. 2,076,308** (Mar. 1, 1993), Y. Okamoto, Y. Hirakouchi, and M. Hagiwara (to Lion Corp.).
36. I. Mathur, *Proceedings of the TAPPI Pulping Conference*, TAPPI Press, Atlanta, Ga., 1991, p. 1015.
37. H. Gottschalk, *Papier (Darmstadt)* **44**(10), 538–540 (1990).
38. M. Liphard, M. Schreck, and K. Hornfeck, in Ref. 36, p. 1031.
39. F. J. Sutman, I. Phillippe, and P. LeBlanc, in Ref. 15, p. 399.
40. S. Dessureault, G. Michaud, S. Tremblay, and C. Massicotte, *Proceedings of the TAPPI Pulping Conference*, TAPPI Press, Altanta, Ga., 1993, p. 875.
41. J. K. Borchardt, D. W. Matalamaki, V. G. Lott, and J. H. Rask, in Ref. 15, p. 449.
42. A. C. Alexander, R. Kurtz, and D. H. McBride, in Ref. 20, p. 197.
43. N. Fetterdy, *Prog. Paper Recyc.* **1**(3), 24 (May 1992).
44. Y. Kotani, *Prog. Paper Recyc.* **3**(3), 110 (Aug. 1993).
45. B. Drehmer and E. Back, *Proceedings of the 3rd Research Forum on Recycling*, Canadian Pulp and Paper Association, Montreal, Canada, Nov. 1995, p. 141.
46. B. M. Balos and J. V. Patterson, in Ref. 20, p. 79.
47. J. E. Angulo, in Ref. 20, p. 229.
48. G. Hanchett, *Prog. Paper Recyc.* **3**(2), 24 (Feb. 1994).
49. C. K. Fallon, in Ref. 8, p. 263.
50. **U.S. Pat. 5,338,401** (Aug. 16, 1994), K. Hristofas, V. L. Magnotta, and R. C. Naddeo (to Air Products and Chemicals, Inc.).
51. M. Doshi, *Prog. Paper Recyc.* **1**(2), 61 (Feb. 1992).
52. C. J. Yu, R. F. DeFoe, and B. R. Crossley, in Ref. 8, p. 451.
53. B. H. Upton, G. A. Krishnagopalan, and S. Abubakr, in Ref. 15, p. 17.
54. G. L. Stanley, *Tappi J.* **79**(1), 37 (Jan. 1996).
55. G. L. Stanley, *Prog. Paper Recyc.* **5**(1), 15 (Nov. 1995).
56. H. Fattah, *Chem. Week* **157**(24), 18 (Dec. 20–27, 1995).
57. *Paper Ind.* **12**(1), 1 (Jan. Feb. 1995).
58. P. N. Williamson, *Tappi J.* **79**(1), 55 (Jan. 1996).
59. *Am. Paper Maker* **25**(12), 42 (Dec. 1991).
60. *Chem. Mark. Rep.*, SR16 (Apr. 26, 1993).
61. A. Loesel, *Chem. Mark. Rep.*, SR 24 (Sept. 28, 1992).
62. *Jpn. Chem. Week*, 6–7 (Feb. 6, 1992).

John K. Borchardt
Shell Chemical Company

RECYCLING, PLASTICS.

See SUPPLEMENT.

RUBBER

In 1839, Charles Goodyear developed a technique for vulcanizing rubber. He blended natural rubber with sulfur and placed it on his wood stove, then noticed that the heat from the stove had cured the rubber into a water-impermeable sheet on one of his failed mailbags (1). In 1844, Goodyear applied for and received a patent for his invention. Because of the shortage of natural rubber at that time, a rubber-reclaiming process, based on steam pressure to devulcanize used rubber, was developed in 1858 (1). The cured ground rubber was subjected to steam pressure for 48 hours. High cost silicone rubber polymers are reclaimed in much the same way in the 1990s, although most synthetic polymers require more complicated techniques.

Rubber recycling has been extended to the use of rubber in asphalt (qv), scrap rubber as fuel, rubber pyrolysis, tire splitting, and other uses. However, the discovery of plastics and oil-extended rubbers, and the drop in crude oil prices (from about \$189 m³ (\$30 bbl) to \$88 m³ (\$14 bbl)), led to a reduction in rubber recycling during the 1970s and 1980s, except for expensive polymers, such as silicones and fluorocarbons (see SILICON COMPOUNDS, SILICONES; ELASTOMERS, SYNTHETIC FLUOROCARBON ELASTOMERS). Pyrolysis of scrap rubber was expensive and markets were limited. It was more expensive to prepare and burn scrap rubber for fuel than to burn natural gas, fuel oil, or coal (qv). High fuel costs and petroleum scarcity in Europe and other parts of the eastern hemisphere have resulted in the use of recycled rubber as fuel and have made it more economical than in the 1970s and 1980s in the United States. With the technological advances of the 1990s, recycled rubber is used more and more in the United States, especially considering the regulations for handling and disposing of worn tires. With fees collected by recyclers, tire recycling is becoming more profitable in the 1990s.

Tires discarded in landfills tend to float on top of the ground; mosquito infestation and illegal tire disposal cause problems which can be alleviated by recycling. Approximately 45% of scrap rubber, primarily as tires, is discarded in landfills (2). Private landfills may charge up to \$3 per tire, and disposal costs at municipal landfills are ca \$0.30–0.60 per tire (3,4), encouraging illegal disposal. As the tire piles grow, rubber recovery becomes more economical (5,6).

In the United States, 48 states have some legislation or regulations dealing with scrap tires; 37 states have enacted specific scrap tire legislation. In 1993, 21 states passed some legislation dealing with scrap tires, many amending laws passed in prior years. The other 16 states regulating scrap tires have broad solid waste laws. These laws and regulations often have many common features which include regulation of the manner of tire disposal in landfills or stockpiles; regulations (including permitting requirements) for scrap tire haulers, processors, and end users; financial responsibility requirements for sites where scrap tires are collected, stored, or processed; provisions for market development, including financial assistance to end use markets or new entrants into the scrap tire business; provisions for financial support for implementation of the state's scrap tire management program; and provisions for clean-up of scrap tire piles.

The tire industry has consistently supported sound scrap tire legislation at both the state and federal level to ensure that scrap tires are managed in environmentally responsible ways. The tire industry has been a consistent participant in the effort to find appropriate markets for scrap tires, and to promote the expansion of those markets.

Some states have restricted the disposal of tires in landfills, eg, Washington, Minnesota, and Oregon. Most states have enacted legislation for a surcharge on the sale of new tires to assist in proper management of scrap tires. Some communities refuse old tires; others require tire splitting or shredding to prevent floating problems in landfills. Fires resulting from the storage of tires have led to regulations governing stockpiles. Mosquito infestation caused by illegal disposal of tires has prompted Saginaw, Michigan to purchase a shredder for landfill disposal.

The rate that used tires are generated is estimated at an annual replacement of 240 million tires. That translates into 350 million passenger tire equivalents (PTE) that require handling each year (one PTE represents a 20-lb (9.1 kg) passenger tire). Sources of used tires include dealers and stores, auto dismantlers, and factory rejects. Recovery alternatives include immediate recycle recovery or stockpiles landfills. Almost 95% of the scrap tires are returned by the consumer to dealers when worn-out tires are replaced by new tires. In 1990, a group called the Scrap Tire Management Council (STMC) was formed as a division of the Rubber Manufacturers Association (RMA). The STMC coordinates with tire companies to address the scrap tire challenge, works on expanding the environmentally and economically sound markets for scrap tires, and is helping to expand market development, information collection and dissemination, research, industry relations, government relations, and encouragement of new technology. The goal of the STMC is the expansion of markets to be able to utilize all scrap tires generated annually (7).

Advances in the technologies of shredding tires, reclaiming rubber, retread equipment, and energy-related projects are helping to resolve the disposal problem. It is estimated that there are more than 700 million whole tires stockpiled throughout the nation (4). Technical and marketing professionals associated with the recycling and disposal industries are paying more attention to tire recycling and energy-generating projects.

Scrap Rubber as Fuel Source

The use of scrap rubber for fuel offers the best alternative for reusing rubber, as fuel costs increase and tire disposal problems become more serious. There are four main markets for scrap rubber fuel, including fuel for cement kilns, electric utilities, pulp and paper mills, and dedicated tire-to-energy plants. Scrap tires used as supplemental fuel by these industries reduce solid waste and air pollution (4). Shredding is not expensive, therefore tires reduced to 2.50-cm chips are most economical. Tires contain more than 90% organic materials and have a heat value of ca 32.6 MJ/kg (ca 14,000 Btu/lb), compared with coal values of 18.6–27.9 MJ/kg (ca 8,000–12,000 Btu/lb). A cyclonic, rotary hearth boiler fired with whole tires was operated by the Goodyear Tire and Rubber Company from 1975 to 1977. It was designed to burn 1400 kg/h and generated 11,300 kg/h of steam (5,8).

In 1991, Goodyear began working with Cadence Environmental Energy (Indiana) to market a whole tire feed system to supplement fuel for cement kilns. The system is used by several cement manufacturers. In 1992, Goodyear furnished tires for a Tennessee Valley Authority (TVA) test burn at a Memphis power plant. The electric utility used tire-derived fuel (TDF) to supplement coal fuel in a cyclone boiler. These tests were successful.

Tire-derived fuel, either as whole tires or processed into chips, is the largest single current and potential market for scrap tires. Tires have a higher energy value than most coals and are cleaner burning. The materials used to construct a tire have positive energy value. With lower emissions than coal, use of TDF has an added bonus of helping reduce air pollution (qv). The final advantage for most fuel users is that TDF can be obtained at a much lower cost per Btu than conventional fuels. From an environmental standpoint, the use of TDF also helps solve a significant solid waste management problem. In 1996, 27 cement kilns, 12 pulp and paper mills, and 8 coal-fired utilities are using TDF as a supplemental fuel. Another 25 cement kilns, 10 pulp and paper mills, and 8 coal-fired utilities are testing or planning to use TDF.

Cement Kilns. The use of scrap tires as a supplemental fuel in cement kilns has had a fairly long history in Germany and Japan, where tires have been used for fuel since the early 1970s. Initially, the scrap tires were purchased by kilns for their fuel value. As the relative cost of other fuels has

declined in Japan, so has the value of scrap tires, and Japanese kilns are collecting a tipping fee for using tires.

In the late 1970s and early 1980s, efforts were made to interest U.S. cement kilns in using scrap tires, either whole or processed, as a supplemental fuel. At least one conference was held on the subject, involving the cement industry, scrap tire processors, and the new tire industry. Although a report was issued by the U.S. Department of Energy (DOE), there was no substantial rush to use tires for fuel. However, some work was done privately with U.S. kilns, and at least a few began testing. By 1990, however, only a few kilns were using tires or planning to do so.

The picture has changed quickly since 1990; there are now 27 U.S. kilns using TDF, either whole or processed, on a production basis, and another 25 are in the permitting or testing process. Several factors caused this change: success of early pioneers caused other kilns to look at tires as a way to reduce energy costs and become more competitive; opposition to the use of other, more environmentally alarming supplemental fuels such as hazardous waste; success of foreign parent companies of many U.S. kilns in using TDF; and growing interest of state scrap tire regulators in finding sound markets coupled with a maturing understanding among regulators of the excellent environmental fit between tires and cement kilns. The Scrap Tire Management Council has been helpful through efforts to expand this market, and by facilitating access to patented Bridgestone technology (2).

If the 25 to 30 kilns in various levels of permitting and testing follow through, nearly one-half of the operating kilns in the United States will be using tires at an average rate of one million tires per year per kiln, or a total of 50 million tires per year. For the next year, the only factor that could slow down this progress is the solid growth in the economy and the expansion of construction activity. The demand for cement was very high for 1994, with many kilns already sold out. Kilns not yet using tires may be reluctant to shut down operations to make the necessary modifications to use tires.

Required air testing is a hurdle any cement kiln encounters as it seeks a permit to use tires as fuel. There is no standard protocol for testing cement kilns using TDF as a supplemental fuel. Each state establishes the testing to be required. The STMC has a research project underway which may remedy this situation. The results of all air emissions testing previously conducted during permitting of cement kilns is being collected and an analysis is being done to understand the emissions which might be affected using tires, and those which have been completely unaffected. The aim is to construct a testing protocol which would deal with only those emissions that might be affected by the use of tires.

Pulp and Paper Mills. The pulp and paper mill market has been one of the principal markets for TDF. Although expansion has been slow, there are 12 facilities using TDF and another 10 in testing or permitting. Often such facilities are in locations where other markets for scrap tires are limited. If the volume of use is such that just one mill in a state, such as Arkansas or Maine, can use more TDF than the entire state can produce annually, the pulp and paper mill focuses on a regional, multistate TDF market. Pulp and paper mills using TDF cover the geographical range of the industry, from Idaho and Oregon, to Wisconsin, Maine, Virginia, Georgia, and Arkansas. A key issue for mills is the need for a substantially dewatered, consistently sized tire chip.

Electricity Generators. The use of TDF as a supplemental fuel in electrical generation is one of the largest current and potential markets for scrap tires. Originally, many observers believed this market would follow the path established by the development of the original Oxford Energy plant (Modesto, California) which is fueled exclusively with scrap tires. Although the original facility, which consumes nearly five million tires per year, was followed by a second facility in Sterling, Connecticut, which consumes nearly 11 million tires per year, efforts by Oxford to build additional facilities in New York and Michigan ran into substantial local opposition. In the meantime, the mainstream electric utility industry began considering the use of TDF to supplement coal in a number of different types of boilers. The tire industry conducted development and testing at in-plant facilities to gauge the feasibility of using tires for energy recovery, and shared the results with electric utilities having similar boilers. The expansion of this market will be through use of tires as a supplemental fuel in existing coal-fired generation stations.

Although Oxford Energy is bankrupt, its two facilities continue to operate, each under separate management. CMS, which initially took over both projects, continues to be involved as the majority owner of the Exeter Energy project (Sterling, Connecticut).

Illinois Power and Waste Recovery, Inc. (WRI), the largest producer of TDF in the United States, entered into an agreement under which WRI will supply TDF to Illinois Power's Baldwin station. Beginning in late 1994, about 3% of the coal was replaced with tires. This requires 70,000 tons of TDF per year, or the equivalent of seven million passenger tires, and represents ~60% of the scrap tires generated in Illinois each year.

Another potentially large user of TDF is the Tennessee Valley Authority. TVA is in the final year of a three-year test of TDF at its Allen station (Memphis, Tennessee). If this facility begins using TDF on a production basis, and if one or two other TVA facilities join in, the potential market for scrap tires would be enormous, consuming some or all of the tires generated in several states.

The use of scrap tires in industrial facilities is also a potential growth market. Perhaps the best known user in this category is Archer-Daniels-Midland Company (Decatur, Illinois). Another industrial user, Flexsys (Sauget, Illinois), uses energy derived from its use of TDF as a supplemental fuel to manufacture chemicals used in the rubber industry.

In 1990, a test using scrap tires (2 × 2 in. TDF) to generate steam for electricity was conducted at the Flexsys plant. The TDF replaced 20% of the plant's coal. Results showed that TDF is an environmentally sound fuel. Particulate emissions were reduced by the lower ash content of TDF, volatile organic compounds (VOC) were reduced because of more efficient burning of TDF compared to coal, and carbon dioxide emissions were reduced because TDF contains half the fixed carbon found in coal. Nitrogen oxide, chlorine emissions, and metals were also reduced, and ferrous metals and dioxins were nondetectable (7).

Industrial Uses. Large industrial facilities, particularly those using cyclone boilers or fluidized-bed boilers, are potential markets. In addition, several vendors of small- and medium-sized industrial energy and steam facilities are marketing units capable of using TDF. As the availability of TDF expands with new producers entering the market, it is hoped that the industrial use of TDF will also expand (7).

In the Lucas tire-burning furnace (Michigan), whole tires are conveyed into an airtight chamber, then onto the outer rim of a rotating hearth. The chamber prevents flashback fires and limits air leaks. An air velocity head of 5.1 cm provides the turbulence necessary for combustion (9). Residues from the burning tires form a char that increases combustion heat loss and tends to clog furnace grates, but the carbon black content reduces staging problems. Improper combustion at the ash removal area of the furnace may prevent burning of carbon black. The Lucas–Goodyear furnace was shut down because of mechanical problems and failure to comply with Michigan's air quality standards (10).

Shredded tire chips have been burned in stoker-fired boilers. Uniroyal fired a 15% mixture of tire chips with coal, and both General Motors and B. F. Goodrich have burned a 10% tire chip mixture with coal (11–13). Tire grinding size reduction problems and delivery costs have stymied projects based on combined tire and coal fuel. Transportation of tire scrap can cost \$0.05/kg, exclusive of grinding costs, thus tire-fired boilers are limited to areas with ample scrap tire supplies, eg, large cities or tire manufacturers. The cost of burning one metric ton of tires per hour in an incinerator was ca \$0.20–0.40 per tire in 1974, which increased to \$0.35–0.70 per tire in 1987 (14).

The Oxford Energy Company used a technology developed by the German Gummi-Mayer Company to incinerate tires and produce electricity. The technology has been used in the Modesto project facility (Westley, California), which generates 14.4 MW of electricity. The project is designed to generate electricity by incinerating whole waste tires on reciprocating stoker grates located beneath two water wall boilers. The boilers produce 55,000 kg/h of steam and power a single General Electric turbine generator. The facility was constructed next to the Filbin Tire Collection Agency, Inc., which administers one of the largest stockpiles of waste tires in the United States (ca 300 × 10³ t); the stockpile increases annually by 27 t. Oxford Energy was also involved in the construction of a New Hampshire tire incineration project, generating 15 MW of electricity; the facility was expected to be completed in late 1988 at a cost of \$35 × 10³. Oxford built a 22-MW electricity generating facility in northeastern Connecticut; construction was completed in 1989 at a cost of \$50 × 10³. The company has projects to incorporate waste produced from manufacturing facilities along with waste tires, thus providing energy to manufacturing facilities as well as local utilities. Approximately 20 × 10³ tires are stockpiled at two sites in New England. These sites, together with the Filbin stockpile, contain at least 480 × 10³ t of tires (53 × 10³ tires) (15).

In 1983, Basic Environmental Engineering, Inc. (Glen Ellyn, Illinois) constructed an incinerator boiler system that burns whole tires at a Firestone plant in Des Moines, Iowa. Here, 20 t of tire are burned daily with 68 t of cardboard, paper, and food to produce about 11,000 kg/h of steam. The project cost \$2.7 × 10⁶. Another incinerator boiler, installed at a Firestone plant at Decatur, Illinois in May 1984, cost \$2.6 × 10³. The incinerator, which has a three-stage combustion control that meets state and U.S. EPA emission levels, burns tire rejects and rubber waste from the manufacturing process and tire dealers. Steel-belted wire from radial tires is handled in a pulse-hearth furnace that mixes waste with combustion air to prevent clogging by the steel wires. Industronics, Inc. (South Windsor, Connecticut) manufactures Consertherm waste to energy systems for incinerating liquids, solids, and semisolids. In one facility, 6000 kg/d of shredded tires are burned, which generate 2200 kg/h of steam at 380 kPa (55 psi) for use in drying tobacco. The Consertherm system

consists of a shredder, conveyor, feed hopper, auger feed, incinerator, ash remover, heat-recovery boiler, and environmental control equipment (scrubber). The incinerators range in capacity from 200 to 400 shredded tires per day. Consertherm incinerators also burn whole tires and a project is underway in Gardner, Massachusetts to process 91,000 kg of whole tires daily to generate 8 MW of electricity (16).

Waste Recovery, Inc. (WRI) shreds tires for fuel at installations in Portland, Oregon; Atlanta, Georgia; Marsilles and Du Po, Illinois; Conshohocla, Pennsylvania; and Houston, Texas. Goodyear owns ~10% of WRI. Firestone burns rejected tires in addition to other solid wastes.

In Japan, scrap tires are used to fuel Portland cement kilns. The process was developed by the Bridgestone Tire Company and has increased greatly in importance since 1980. The Saitama plant (Nihon Co., Ltd) burns 140,000 tires, as well as oil, per month to produce 258,300 t of cement (qv) per month (17). Cemenergy (Loomis, California) and Emanuel Tire Company (Baltimore, Maryland) cement companies and paper mills burn scrap tires to produce process energy (18,19). It has been predicted that 238 million passenger and truck tires will be consumed by the energy market annually by 1997 (20).

Pyrolysis

Scrap tire pyrolysis has been the subject of several studies by rubber, oil, and carbon black industries. The Tosco II process pyrolysis study sought to develop equipment to maximize pyrolytic filler production and quality (Fig. 1) (21–24). Chopped tires are fed into a rotary drum containing hot ceramic balls at 480°C in a reducing atmosphere. The rubber pyrolyzes and forms a solid residue, oil vapor, and off-gases; the condensed oil separates in a fractionator and the gas is used to heat the ceramic balls. A trommel screen separates the fine solid residue from the ceramic balls, and this residue is pelletized after steel, fiber glass, and other contaminants have been removed. The off-gas is a combination of ethylene, propylene, and butylene; the oil contains about 1% sulfur and can be substituted directly for fuel oil. Higher temperatures produce more gas and less liquid. The pilot-plant process was designed to handle 13.5 t d of tires, generating 0.5–0.6 m³ (3–4 bbl) of oil, 1270–1540 kg of pyrolytic filler, 190–220 kg of steel, and 154–176 kg of fiber glass. The Tosco II project was completed, but additional projects were discouraged, because the U.S. carbon black industry already had excess capacity. Furthermore, the types of pyrolysis products, eg, ZnO and other inorganic materials and glass fiber from old tires, are unpredictable. However, pyrolytic fillers must be carefully chosen to achieve specific properties in compounded rubbers, thus the residue mixture from nonvacuum pyrolysis is useful only as a low grade filler and cannot be used as a substitute carbon black source. Since 1992 there has been an undercapacity situation for carbon black and tire pyrolysis is being considered to produce pyrolytic filler.

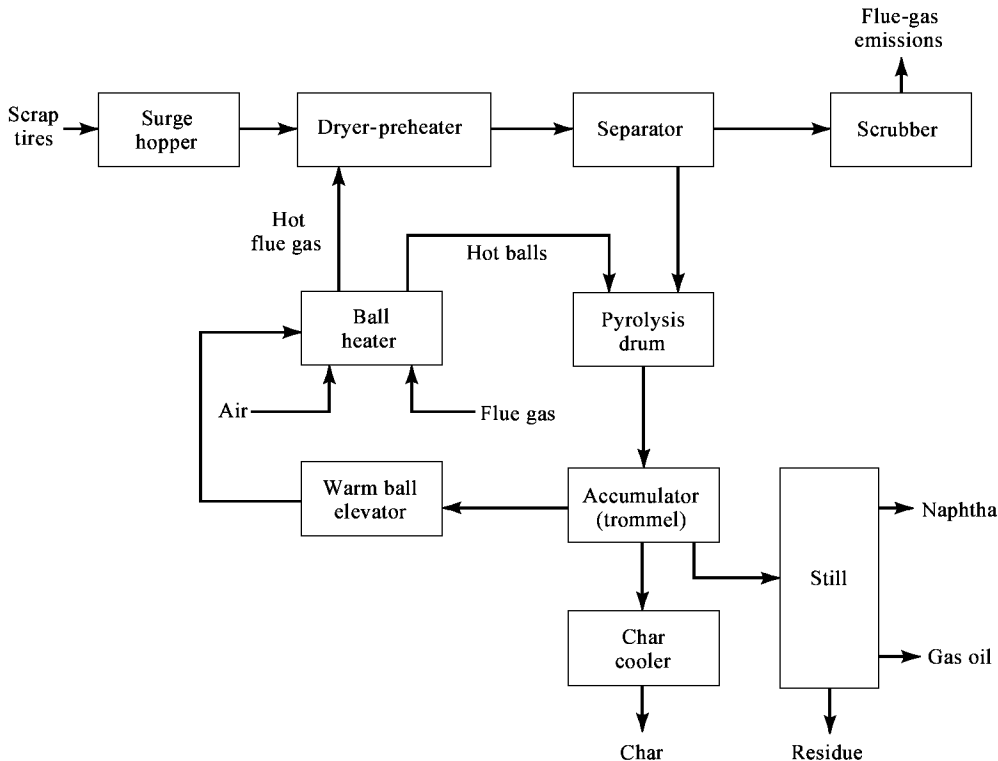


Fig. 1. The Tosco II process.

Foster-Wheeler has two pyrolysis plants operating in Germany; other European countries and Japan also pyrolyze scrap rubber. Tyrolysis built a 122 t d pyrolysis plant in the United Kingdom (24), employing vertical cross-bow reactors (25). Intennco (Houston, Texas) shut down a 40 t d scrap tire pilot plant. In the Intennco process, the shredded tires were pyrolyzed by two reactors in series by indirect heat at ca 540°C in a reducing atmosphere. The process yielded ca 0.5 m³ (3 bbl) of aromatic (30–40% light) oil per ton of scrap rubber. Contaminants in the pyrolytic char were removed by several proprietary processing steps (26). In the United States, tire pyrolysis has been studied as a method for recovering pyrolytic filler; however, production of virgin carbon black from petroleum oils is less expensive, easier to control, and gives better quality.

Whole tires have been pyrolyzed in an experimental semifluidized-bed reactor (27). The tires are pyrolyzed on a grate that tilts to discharge the steel belt and bead wire. The pyrolysis of whole tires eliminates shredding and grinding costs.

American Tire Reclamation, Inc. (Detroit, Michigan) plans to construct a commercial-size plant to supply a refined version of carbon black and oil produced by pyrolysis of tires. The pyrolysis product improves aging and reduces rutting when added to asphalt (qv) (28).

Nippon Zeon estimated that the break-even cost of its tire pyrolysis pilot plant was \$0.25 per tire (29,30). One study indicates that pyrolysis of tires and other polymers should be considered as a means for disposing of scrap within environmental constraints. A plant processing 81,000 t yr of scrap could be profitable, based on sales of reclaimed products (31).

Conrad Industries, Inc. (Centralia, Washington) and Clean Air Products Company (Portland, Oregon) have jointly built a tire pyrolysis demonstration machine which allows recovery of combustible gases, oils, and other by-products. The equipment can also handle other carbonaceous material. It is designed to process 0.9 t h of tires; the entire system is estimated to cost about \$2.3 × 10⁶. The feedstock consists of 5-cm tires chips which produce pyrolytic filler, a vapor gas yielding 11.5 kJ m³ (1000 Btu ft³), and medium and light oils yielding about 42 MJ kg (18,000 Btu lb) (32).

Energy Conservation Corporation (Chadds Ford, Pennsylvania) has constructed a tire pyrolysis cogeneration demonstration facility in Derry, New Hampshire. Waste tires are fed into a continuous-line pyrolysis system to produce 5 MW of electricity at a \$6.8 million coal-generation facility. The demonstration facility will process 5000 tires per day and produce 13,600 kg of pyrolytic filler, clean steel, oil, and noncondensable gases. In this process, 10-cm shredded tire chips are fed to a retort chamber where they are vaporized at 700°C. The vapors from the retort are passed through a condenser and the solids are transferred to a second retort where they are segregated and conveyed across a water bath to individual pyrolytic filler and steel storage bins. The gases from the process are used as an energy source. The pyrolysis cogeneration facility was not operating as of early 1996.

Enerco, Inc. (Yardley, Pennsylvania) has a 600 tire d demonstration pyrolysis plant located in Indiana, Pennsylvania. The facility operated 8 h d, 5 d wk for six months. The process involves pyrolysis in a 5.4 t d batch-operated retort chamber. The heated tires are broken down to crude oil, noncondensable gases, pyrolytic filter, steel (qv), and fabric waste. In this process, hot gases are fed directly to the rubber rather than using indirect heating as in most other pyrolyses. The pyrolysis plant was not operating as of early 1996.

Kutrieb Corporation (Chetek, Wisconsin) operates a pyrolator process for converting tires into oil, pyrolytic filler, gas, and steel. Nu-Tech (Bensenville, Illinois) employs the Pyro-Matic resource recovery system for tire pyrolysis, which consists of a shredding operation, storage hopper, char-collection chambers, furnace box with a 61-cm reactor chamber, material-feed conveyor, control-feed inlet, and oil collection system. It is rated to produce 272.5 L oil and 363 kg carbon black from 907 kg of shredded tires. TecSon Corporation (Janesville, Wisconsin) has a Pyro-Mass recovery system that pyrolyzes chopped tire particles into char, oil, and gas. The system can process up to 1000 kg h and produce 1.25 MW h (16).

American Ligurian, Inc. (Stanford, Connecticut) is marketing a pyrolysis process developed in Italy, which generates steam for hot water, air heating, dryers, kilns, and similar installations. A modular plant produces 8000 kg h of steam from 1 tih of tires. The pyrolysis process produces 0.9 t of fuel oil, 270 t of steel, and 54 t of ash annually. Gas emissions meet the strictest environmental standards (16).

Other techniques include oxidative, steam atmosphere (33), and molten salt (34) pyrolyses. In a partial-air atmosphere, rubber pyrolysis is an exothermic reaction. The reaction rate and ratio of pyrolytic filler to oil products are controlled by the oxygen flow rate. Pyrolysis in a steam atmosphere gives a cleaner char with a greater surface area than char pyrolyzed in an inert atmosphere; however, the physical properties of the cured compounded rubber are inferior. Because of the greater surface area, this pyrolytic filler could be used as activated carbon, but production costs are prohibitive. Molten salt baths produce pyrolyzed char and oil products from tire chips. The product characteristics and quantities depend on the salt used. Recovery of char from the molten salt is difficult.

Depolymerized Scrap Rubber

In the depolymerized scrap rubber (DSR) experimental process, ground scrap rubber tires produce a carbon black dispersion in oil (35). Initially, aromatic oils are blended with the tire crumb, and the mixture is heated at 250–275°C in an autoclave for 12–24 h. The oil acts as a heat-transfer medium and swelling agent, and the heat and oil cause the rubber to depolymerize. As more DSR is produced and rubber is added, less aromatic oil is needed, and eventually virtually 100° of the oil is replaced by DSR. The DSR reduces thermal oxidation of polymers and increases the tack of uncured rubber (36,37). Depolymerized scrap rubber has a heat value of 40 MJ kg (17,200 Btu lb) and is blended with No. 2 fuel oil as fuel extender (38).

Rubber, Asphalt Modification

Ground recycled rubber has been developed as a modifier for asphalt paving materials since the 1950s in both the United States and Europe. The original developmental work was undertaken in an effort to improve the performance of asphalt paving materials, not to find a method to dispose of scrap tires, thus many different additives and modifiers have been tested with some becoming accepted practices. The utilization of various polymers has been investigated (28). Two basic processes, wet and dry, have been developed, differing in the point at which the ground rubber is introduced.

In the wet process, also called the McDonald or Arizona process after its developer and the location in which it was developed, respectively, ground rubber is prereacted with the hot liquid asphalt cement for a specified time, depending on the size of the rubber particles. During this process, the rubber particles react with and change the asphalt cement. The rubber-modified asphalt can then be used in one of several ways: as a crack sealant in pavement repair or with an aggregate chip as a cape sealant to effect temporary repair; or, in the rehabilitation of cracked pavements to prevent or reduce subsequent reflexive cracking, it can be sprayed over cracked pavements as a stress-absorbing membrane or with a thin layer of asphalt concrete as a stress-absorbing membrane interliner (SAMI) and then covered with a new friction course. Rubber-modified asphalt cement can also be mixed with the normal aggregate components to produce a rubber-modified asphalt concrete. This material can be placed in the usual manner, with some limited modifications, and used in both base and friction courses.

The dry process was originally developed in Sweden and involves using slightly larger rubber particles mixed with the dry aggregate materials. The aggregates, which must be gap-graded to allow room for the rubber particles, are mixed with normal asphalt cement to form a modified asphalt concrete. This material can be placed in the usual manner, with limited modifications, to the handling procedure (39).

The United States generates ca 250 × 10³ scrap tires per year. In 1977, ca 52 t of reclaimed rubber and 14,400 t of crumb rubber were produced from these tires (40). In 1980, ca 4050 t of reclaimed and crumb rubber were used in asphalt–rubber compounds, which is less than 5° of the recycled rubber. Except for a reduction in reclaimed rubber, the 1986 production rates were higher; an estimated 25,200 t of rubber were used in asphalt.

The largest potential market for ground rubber is as an additive to hot-mix asphalt concrete used as a paving material. While this use for scrap tires has been in various stages of research, development, and implementation on a small scale for a number of years, it came to the forefront of scrap tire uses with the passage of the Intermodal Surface Transportation Efficiency Act of 1991 (ISTEA). Section 1038 of the ISTEA requires all states to use scrap tire-derived ground or crumb rubber in asphalt paving in federally funded highway construction as of 1994, starting at 5° of paving and escalating to 20° by 1997. The issue has been controversial.

If the requirements of Section 1038 of the Act are fully implemented, it is estimated that 70 million tires would be required on an annual basis. If states found reason to exceed the law's minimum use of scrap tire-derived ground rubber in asphalt paving and use more on federal aid construction, or if local, county, and state governments and private parties started using rubber-modified asphalt in nonfederal aid construction, the totals could go much higher.

Some organizations are against the use of rubber-modified asphalt and are working to reduce or eliminate the requirement. These organizations were successful in having the U.S. Congress enact a one-year moratorium on the Section 1038 requirement in 1994 and 1995. Efforts are underway to eliminate or substantially revise this provision. The ground rubber industry, which would develop the scrap tire product to meet the demands for rubber-modified asphalt, is awaiting the outcome.

The existing ground rubber producers have expanded their capacity to meet anticipated demands. Potential new entrants have been forced to decide whether to go ahead with investment plans, or await a final Congressional verdict. State transportation department groups that lead the anti-Section 1038 faction are also deciding whether to proceed with test sections and other preparations for implementing the requirement.

In Arizona, California, and Florida, the use of rubber-modified asphalt will likely expand whatever the outcome with ISTEA, because of less stress cracking. The use of rubber-modified asphalt may increase in other states if they follow California's lead and allow reduced-lift thickness when using rubber-modified asphalt in pavement rehabilitation (7). The Federal Highway Administration has undertaken studies to provide guidance to those states that must comply with Section 1038.

Asphalt–rubber is mixed and applied to roadways by several techniques. In one method, rubber and asphalt are mixed at ca 175–220°C for one to two hours. The hot mixture is applied to the roadway and covered with a layer of stone chips to form a chip seal. The rubber crumb consists of scrap tires ground into particles less than 2 mm in diameter. Rubber-modified asphalt is also used for waterproofing membranes, crack-and-joint sealers, hot-mix binders, and roofing materials. The rubber improves asphalt ductility and increases its softening point. The aggregate adhesive bond is stronger, and the asphalt lasts longer. Production of rubber-modified asphalt has increased from 405 t in 1970 to 27,000 t in 1980 (41). Typically, about 2 t of rubber is used for 1 km of roadway. If it is assumed that asphalt–rubber contains ca 25° rubber and 75° asphalt, the potential demand for scrap rubber would be ca 40,500 t yr, or ca 2° of the amount available.

The use of asphalt–rubber in roadways was encouraged by the Federal Highway Administration in Arizona during the early 1970s (16). Uniroyal and other companies involved in rubber reclaiming during the late 1960s and early 1970s conducted research to develop asphalt–rubber formulations suitable for roads, joint sealer, and recreational surfaces (qv), including running tracks, tennis courts, and playgrounds.

The Arizona Department of Transportation is increasing application of a three-layer asphalt–rubber system based on test results of heavily traveled highways. Other states testing asphalt–rubber include New York, Massachusetts, New Jersey, Louisiana, and Alabama. The U.S. EPA is proposing

guidelines to encourage this development.

All Seasons Surfacing (Bellevue, Washington) was a licensee of Swedish technology for processing waste rubber into asphalt-paving materials called Plusride. Although no longer in business, All Seasons developed a process for paving material formed by removing portions of aggregate from asphalt and replacing it with rubber granules. In this process, 3% of 0.5-cm rubber tire chips and 20% powdered rubber dust are used. The *Rebounder*, a publication of the Rubber Pavements Association (Washington, D.C.), reported that 540 t of Plusride asphalt–rubber composition was utilized in New Jersey. The company was also exploring markets in Massachusetts and Louisiana. The material performed well in locations with severe temperature differentials; it was resistant to studded tires and chain wear and saved pavement because less sanding and salting were needed in icy conditions. Plusride was used to pave roadways destroyed by the eruption of Mt. St. Helens (Washington) in the late 1980s. In spite of the scale of operation, performance of Plusride compared well with normal usage and no maintenance has been required (16). For 1986, there was a 20–25% increase in the use of rubber for highway applications.

Asphalt-Rubber Systems (Warwick, Rhode Island) markets, designs, and develops specifications for asphalt–rubber paving projects for highways and streets. The company has paved 48 km of highway in Worcester, Massachusetts. Some sections of the road paved with asphalt–rubber membranes are more resilient than surfaces without asphalt–rubber (16).

Ground rubber used as an asphalt modifier has been an accepted material in Arizona, California, Florida, and Texas. Greater acceptance will require that state highway departments become familiar with rubber modification of asphalt paving materials and gain confidence in its capabilities. Efforts are underway to ensure greater technology transfer and information dissemination on asphalt–rubber issues. The National Cooperative Highway Research Program of the National Research Council has established the CRM project 20-7 to provide an information network for anyone interested in crumb rubber modifiers. It is administered by the Technology Transfer Center at the College of Engineering, University of Nevada, Reno (42).

Cryogenic Pulverizing and Mechanical Tire Shredding

Tires must be pulverized or shredded before they can be reclaimed by devulcanization or used in asphalt and other recycling processes. The tires are mechanically ground, sometimes using cryogenic or solvent-swelling techniques to enhance grinding efficiency. In one process, a polar solvent is used to swell the rubber, followed by shearing to reduce particle size (43). Ground tire crumb rubber is commonly referred to as rubber reclaim, even though the rubber has not been devulcanized.

Cryogenics (qv) in conjunction with mechanical action has been used to make crumb rubber. Liquid nitrogen or other cryogenic fluids cool the rubber below the glass-transition temperature, and the brittle rubber is pulverized in a grinding mill. A small cryogenic system can be installed at the site to integrate the scrap rubber crumb into the compound mixing process. Cryogenic grinding of different rubber types, related costs, and rubber compound properties is available (40–44). Cryogenic grinding costs are \$0.20–0.40 /kg, depending on the desired particle size and type of rubber; harder rubber is easier to grind. Ground rubber scrap can be devulcanized, pyrolyzed, or recycled directly into the rubber compound. Ground rubber is also added to plastics (45).

Air Products and Chemicals, Inc. (Allentown, Pennsylvania) developed a cryogenic process for grinding scrap tires in the mid-1960s. Midwest Elastomers installed a cryogenic system in 1980 to powder tire peels in its Wapakoneta, Ohio, plant. The equipment and materials for cryogenic processing can be expensive and have slowed expansion. Cryogenic grinding can require proximity to an air processing facility to make the use of liquid nitrogen economical.

In another grinding process, tires are mechanically ground with a two-roll, grooved rubber mill. The two-mill rolls turn at a ratio of ca 1:3, providing the shearing action necessary to rip the tire apart. The rubber chunks are screened and the larger material is recycled until the desired size is reached. Bead wire is removed by hand or with magnets. For most applications, eg, devulcanization or pyrolysis, crumb rubber particles smaller than 1.19 mm (16 mesh) are desired, and several milling steps are required. Tire fiber is removed in intermediate operations with hammer mills, reel beaters, and air tables that blow a steady stream of air across the rubber, separating the fiber. Tire grinding is highly energy-intensive and tire slitters may be used to cut the tire initially.

The morphology of ground rubber particles plays an important role in the ability to incorporate them into new compounds. Ambiently ground particles tend to have a more heterogeneous surface than cryogenically ground particles, the surfaces of which are more regular and cube-shaped. In practice, the adhesion of ground rubber in new compounds is a mechanical process. As a result, ambient ground materials appear to incorporate more easily into new compounds. In order to increase the reincorporation of cryogenically produced materials, work has been done to modify the surface of these particles to make them more reactive when incorporated into new compounds. Such a process has been developed and has received assistance from the U.S. Department of Energy to apply it to commercial applications (46). Given the higher cost associated with this surface modification process, a principal area of investigation is the use of this material in compounds using high value virgin elastomers (42).

In 1985, the Emanuel Tire Company in Baltimore processed more than 3 × 10⁶ tires into chips, which are mostly sold to pulp and paper mills as a supplemental fuel; the remainder is sold to reclaiming facilities or landfilled. Only 20% of passenger tires are suitable for recapping. Nonrecappable tires are shredded into 5-cm chips. The Emanuel Tire operation is capable of reducing the 5-cm chips to smaller sizes. Shredded waste tire chips can be granulated into very fine wire and fabric-free rubber particles.

Rubber Research Elastomerics, Inc. (RRE) plans to operate a \$2.3 × 10⁶ tire processing plant in Babbitt, Minnesota. The plant will generate 7200 t/yr of finely ground foam rubber for feedstocks in its tire-recycling process. A liquid polymer is blended with finely ground crumb rubber to form reclaimed rubber products for use in molded products such as tires and polypropylene elastomers (see OLEFIN POLYMERS, POLYPROPYLENE). The Babbitt facility processes tires without charging a tipping fee. RRE plans to use tires from an existing pile in St. Louis County, Minnesota.

Atlas Fox (Bloomfield, Connecticut) uses a National Standard stationary shredder unit to shred tires into strips 8 × 15 cm. Atlas charges \$35 /t for tires delivered to their facility and collects tires throughout New England. The company hopes to supply the strips for use as fuel in a cogeneration system in the Hartford, Connecticut area. Cemenergy (Redding, California) charges ~ \$0.75 per passenger tire and ~ \$3.75 per truck tire (tipping fee). Louisiana Pacific (Antioch, California) burns 45–67.5 t/d of tires with wood chips in a wood waste boiler to generate steam for electricity for a lumber mill. Cemenergy burns tire chips in a 10% mixture with coal. The steel-belted portion in the chips is integrated into the cement as iron oxide. Considering that coal–sulfur emissions are higher than those from tires, the total emissions of sulfur are not increased by using a tire–coal mix. Cemenergy is investigating a market for using ~ 180 t/d of tire chips to generate power. Their shredders operate on the same principle as those designed by Waste Recovery, Inc.

Other waste rubber products, such as production scrap from the manufacture of molded and extruded products, hose and belt manufacturers, or any other rubber products manufacturing operations, can also be reduced to ground rubber. A tire contains several types of rubber, and different rubber compounds. As a result, the chemical properties of ground rubber produced from whole tires can vary. In addition, different manufacturers use different compounds, and compounding recipes differ with the type of tire, adding to the variability possible with ground tire rubber. Production scrap or off-specification products from other nontire rubber manufacturing processes may contain only a single rubber compound. When this material is ground, it may present more uniform characteristics and be more easily utilized, including in the manufacture of the original product from which it came.

There are several uses for ground rubber: compounding rubber stocks for use in production of various molded rubber products, combining with higher cost elastomers to reduce material costs without affecting product performance, and mixing with other materials, both recycled and virgin, to manufacture products not traditionally made from rubber. Experimentally, ground rubber has been used as the basis of materials for remediation of oil spills.

Ground rubber is used in the manufacture of many products, including friction materials, floor mats and relief pads, new tire manufacture, and molded rubber products. In new tire manufacture, ground rubber is most commonly used as a processing aid. Small quantities of ground butyl rubber are used in compounding the tire inner liner to reduce air entrapment during curing. In addition, small quantities of ground rubber are used as filler materials in various tire components where its use does not compromise the performance required of the compound.

Floor mats, carpet backing, and relief mats have been traditional markets for ground rubber where the material adds resiliency at a lower cost compared to virgin compounds. Certain molded products have also been markets for ground rubber. An interesting use is the manufacture of lawn soaker

hoses from ground rubber and polyurethane binder.

The key to expanding product markets for additional recycled ground rubber is customer acceptance. The auto industry is one of the largest customers of molded rubber products. Until recently, most auto industry specifications for molded rubber parts specifically prohibited the use of recycled rubber. Increasingly, the auto industry is seeking greater use of recycled materials, including recycled rubber, in the parts it purchases. An example is the announcement by Ford that it is testing the use of a compound containing ground recycled rubber to manufacture brake pedal pads, and the Ford–Michelin project (announced in June 1995) to achieve 10% recycled content in new original equipment (OE) tires.

The auto industry has established a cooperative program in the field, the Vehicle Recycling Partnership (VRP). In addition, some auto manufacturers have announced specific projects that include recycled rubber. Chrysler has announced that it is already using a rubber shield that incorporates crumb rubber derived from scrap tires. Auto manufacturing companies have indicated a desire to expand the use of recycled post-consumer scrap rubber in their products.

Another need has been a technical forum in which rubber industry chemists and engineers can share information and insights into the problems raised by the expanded use of recycled ground rubber. This need is being filled by the Topical Discussion Group on Rubber Recycling, organized as part of the rubber division of the American Chemical Society. This group sponsored its first symposium in May 1995 as part of the rubber division's semiannual meeting. The Topical Discussion Group brings together producers of ground recycled rubber, with compounding and engineering personnel from rubber products manufacturers and representatives from customer industries to share information and identify needed activity.

There has been considerable interest in ground rubber as a raw material, with many experimenters and entrepreneurs attempting to develop totally new uses for the material, including using ground rubber and recycled plastics to produce a lumber substitute. Products produced in this manner have been fence posts and flooring for livestock trailers. Other investigators are experimenting with ground rubber as an additive to concrete to give greater resistance to earthquake damage.

Reclaiming

Originally only solid rubber scrap was reclaimed, but with the advent of pneumatic tires and fiber-reinforced rubber, methods for removing the fiber had to be developed. Although the reclaiming processes are of limited use in the 1990s recycling market, modified reclaiming processes may become more prevalent as recycling increases.

Acids and alkalis were used to decompose the fiber to cellulose. The alkali digester process, developed in 1899, is still used. Fiber glass reinforcement must be removed mechanically before the rubber can be reclaimed. A highly efficient method involves hammer mills and reel beaters to separate the fiber from the rubber; an air current subsequently drives off the fiber.

Until World War II only natural rubber was used in tires and reclaiming was simple. Reclaiming plasticizers penetrated the natural rubber evenly, bond cleavage was uniform, and reclaim products were soft and consistent. However, these plasticizers do not penetrate synthetic rubber particles as easily, and synthetic rubbers harden under heat and pressure. Furthermore, they are less susceptible to oxidation than natural rubber. Scrap tires contain a mixture of synthetic and natural rubbers which must be ground to a finer particle size for better oil absorption, and stronger chemical reclaiming agents must be used. Alkylphenol sulfides, aromatic amines, chlorinated mercaptans, and other strong plasticizing agents are absorbed evenly into the rubber particles, resulting in bond cleavage and a uniform smooth reclaim (Table 1).

Table 1. Oils and Processing Aids for Reclaiming Rubber

Material	Composition	Action
Paraflux	saturated plasticizer, poly-merized hydrocarbon	plasticizer
tall oil pitch		
Tarene ^a	mixed terpenes, a pine-tar product	extender; plasticizer aids in dispersion of fillers; softens product and improves processing
Solvenol	monocyclic terpenes	processing aid; nonstaining reclaim oil and solvent; swells and penetrates rubber; dissolves and disperses heavy oils
Pitt Consol 500	aryl disulfides in petroleum oil	peptizer; shortens devulcan-izing time by oxidizing rubber bonds
300' mineral rubber	bituminous petroleum product	extender, plasticizer, and softener; aids in processing
resorcinol	resorcinol	reclaim
		tire formulations; promotes adhesion of rubber to cord; reduces stain in whitewall tires when used with formaldehyde
Corray 40	cycloparaffinic hydrocarbon	processing aid; softens, swells, and smooths reclaim
carbon black		reinforcing filler
Triton $\times 100^b$	alkylaryl polyether alcohol	emulsifier; aids in dispersing oils over rubber particles
clay	aluminum silicate	inert filler; hardens and stiffens reclaim
limestone whiting	calcium carbonate	inert filler; hardens and stiffens reclaim
liquid caustic ^c	sodium hydroxide	devulcanizing agent; removes free sulfur and acts as curing aid

^a Trademark of Harwick.

^b Trademark of Rohm and Haas.

^c 30% NaOH solution.

The three rubber-reclaiming processes are digester, heater or pan, and reclaimator processes (47). Goodyear has devulcanized and reclaimed scraprubber with the help of microwaves (48). Tires are most commonly reclaimed by digesting (Fig. 2). Two-roll mills or other grinding devices reduce whole tires to uniform particles, and fiber is mechanically separated from the rubber with hammer mills, blown into collectors, and baled. Metal chlorides may also be used to reduce tire fiber chemically during digesting. Oils and processing aids are blended with the crumb rubber in ribbon blenders or similar mixers and are transferred to a digester, a steam-pressurized tank equipped with horizontal mixing paddles. The blend is mixed continuously at steam pressures of 1.01–1.70 MPa (10–17 atm) for four to six hours. The pressurized digester batch is forced into a blowdown tank, washed, and dried. Compounding ingredients, eg, carbon black and clay, are added to modify and maintain certain properties required for specific applications. Metal and other contaminants are strained from the digested rubber by extruders. High friction refining mills provide the smoothness and physical properties needed in the final product. The reclaim is baled, extruded into pellets, or formed into slabs for shipment. A typical reclaim formulation is shown in Table 2.

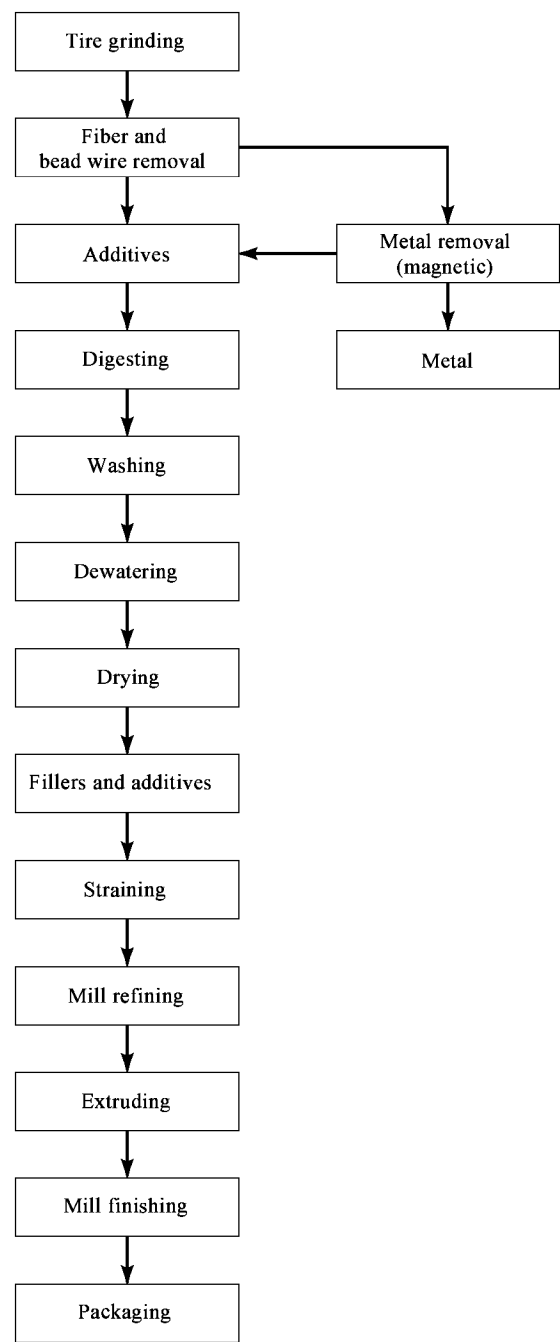


Fig. 2. The reclaim digester process.

Table 2. Formulation for Reclaiming Whole SBR Tires^a

Ingredient	Parts, phr
SBR ^b	100.0
Paraflux	5.4
tall oil pitch	6.0
Solvenol	2.0
Pitt Consol 500	2.0
300' mineral rubber	1.7
carbon black	4.0
resorcinol	0.06
Triton × 100	0.06

^a Digested for 5 h at a steam pressure of 1.01 MPa (10 atm).
^b Styrene–butadiene rubber reduced to <1.19-mm (16-mesh) particles.

Innertubes of butyl and natural rubber and other fiber-free scrap rubbers are reclaimed by the heater or pan process. Brass tube fittings and other metals are removed from the scrap which is mechanically ground, mixed with necessary processing aids (see Table 1), loaded into pans or devulcanizing boats, and autoclaved at steam pressures of 1.01–1.42 MPa (10–14 atm) for three to eight hours. The product is refined by milling, extruded, and milled again in much the same way as in the digester process. A typical reclaiming formulation is shown in Table 3.

Table 3. Formulation for Reclaiming Butyl Rubbers^a

Ingredient	Parts, phr
IIRh ^b	100
Corray 40	2

clay	3
water	2

^a Devulcanized for 5 h at a steam pressure of 1.01 MPa (10 atm).

^b Butyl tubes reduced to 6.25-mm particles.

In the Reclaimator, a high pressure extruder, fiber-free rubber is heated to 175–205°C with oils and other ingredients. High pressure and shear between the rubber mixture and the extruder barrel walls effectively devulcanize the mixture in one to three minutes. In the Lancaster-Banbury method, high temperature, pressure, and shear are applied to the rubber in a batch process that is otherwise similar to the Reclaimator process. In another high pressure process, scrap rubber is devulcanized at 5.5–6.9 MPa (54–68 atm) for ca five minutes. The product is milled, baled, or pelletized as in other processes.

Natural rubber (NR) scrap is reclaimed with solvent naphthas, terpenes, dipentenenes, and resins that swell, tackify, and promote bond cleavage (see RUBBER, NATURAL). Tires, mainly synthetic styrene–butadiene rubber (qv) (SBR), are less susceptible to oxidation (see BUTADIENE). Aryl disulfides, phenyl disulfides, high molecular weight mercaptans, and other sulfur-containing chemicals are used to swell and lubricate the bonds and devulcanize the rubber by chemical oxidation and bond cleavage. The oils control the properties of the product, soften the rubber, and increase the elongation. These properties are also affected by reinforcing agents and fillers (qv), eg, carbon black, aluminum silicate, and calcium carbonate. Nonreinforcing fillers reduce the tack of devulcanized rubber and facilitate handling. They abrade vulcanized particles during the final milling, resulting in a smooth texture. Reinforcing fillers increase the tensile strength of the reclaimed material. Typical reclaiming formulations are shown in Table 4.

Table 4. Formulation for Reclaiming Natural Rubber

Ingredient	Parts, phr
NR black ^b Tarene limestone whiting liquid caustic ^c water	Carbon black ^a
	100
	3
	10
	2
	2
NR neutral ^e Corray 40 clay liquid caustic ^c water	Neutral rubbers ^d
	100
	3
	2
	2
	2

^a Devulcanized for 5 h at a steam pressure of 1.01 MPa (10 atm).

^b Black NR tubes reduced to 6.25-mm size particles.

^c 30% NaOH.

^d Devulcanized for 5 h at a steam pressure of 1.4 MPa (14 atm).

^e NR elastic bands and thread reduced to 6.25-mm size particles.

Tires, natural rubber tubes, and butyl tubes are the main sources of scrap and reclaim (see ELASTOMERS, SYNTHETIC POLYISOPRENE). Specialty reclaim materials are made from scrap silicone, chloroprene (CR), nitrile– butadiene (NBR), and ethylene–propylene–diene–terpolymer (EPDM) rubber scraps (see ELASTOMERS, SYNTHETIC POLYCHLOROPRENE; ELASTOMERS, SYNTHETIC ETHYLENE PROPYLENE DIENE RUBBER). Tires, hoses, belts, molded and extruded goods, and asphalt products consume ca 80% of the reclaimed rubber manufactured. Typical properties of reclaimed rubbers are shown in Table 5.

Table 5. Properties of Reclaimed Rubbers

Polymer	Specific gravity	Mooney viscosity	Tensile strength, MPa ^a	Elongation, %	Ash, %	Acetone extract, %
NR ^b						
neutral	1.19	45	9.0	500	29	9
black	1.19	25	8.6	400	13	11
SBR, whole tire ^b	1.18	50	6.9	280	7	22
IIR, tubes ^c	1.17	65	9.0	540	7	

^a To convert MPa to psi, multiply by 145.

^b Rubber Reclaimers Association (RRA) cure test recipe for mixed elastomer and NR reclaim; cured 20 min at 141°C.

^c RRA cure test recipe for IIR reclaim, cured 30 min at 160°C. IIR is isobutylene–isoprene rubber (butyl rubber).

Until the 1960s, reclaimed rubber was an important raw material in molded and extruded rubber products, eg, tires, rubber mats, and hard rubber battery cases. With the advent of vinyl, other plastics, and less expensive oil-extended synthetic polymers, reclaimed rubber sales stabilized and decreased. In 1973, the oil embargo and rising energy costs increased costs of the energy-intensive rubber reclaiming process to the point where they matched virgin polymer costs. Increased radial tire production required crack resistance that could not be provided by reclaimed rubber compounds (46).

Civil Engineering Market

The civil engineering market for scrap tires encompasses several distinct uses. Whole tires have been used to construct retaining walls and crash barriers. One publicized use is the construction of houses and at least one motel (7). Whole tires have been used in erosion control, and to construct breakwaters and artificial reefs.

Jai Tire has patented a process that, upon adding crumb rubber to soil, improves turf by providing aeration and reducing compaction (49). Superior Environmental Products, Inc. has developed a material that contains 40% of scrap tire rubber. The material can control chemical, oil, and gasoline spills (50).

A broader use of scrap tires as an engineering material is in the formation of chipped or shredded scrap tires as a fill material, especially useful where

light weight is needed. Notable projects include the fill over an underground garage in Minneapolis and the use of shredded tire fill to construct a ramp to an interstate highway where the ramp passes over a closed landfill. The lighter weight of the tire fill is expected to reduce settling as the landfill matures. Research is also being conducted in the use of shredded tires to backfill retaining walls, and as an insulating material under roadbeds in cold climates.

Expansion of the civil engineering markets depends on the availability of standard engineering data for shredded tire material; fortunately, several projects are being undertaken by university-based researchers who are generating this data. Another limiting issue is concern about potential leachates that might result from any of these uses; again, data is being developed which so far has demonstrated no difficulties with tires in these uses.

Several state highway departments have undertaken road construction-related fill projects using tire chips, including Oregon, Colorado, North Carolina, Virginia, Maine, Minnesota, and Wisconsin. Minnesota has achieved considerable private use of tire chips as fill material, as has Vermont. Several states have approved the use of tire chips in private septic systems as backfill for leach fields, including Iowa, Georgia, and South Carolina. Whole used tires are used for a thick flow zone for leachate and as a protective barrier at the Pen Argyl landfill in Pennsylvania (51).

Small tire chips have also been utilized as a soil amendment to improve athletic playing fields (see RECREATIONAL SURFACES). A patented process marketed under the trade name Rebound (Jai Tire) combines crumb rubber from scrap tires with composted organic material to reduce soil compaction, resulting in better athletic playing surfaces (52). Installations have been made in Florida, California, Colorado, Hawaii, Maryland, Michigan, Missouri, Nevada, Virginia, and Wisconsin.

The Americans with Disabilities Act (ADA) has opened another market for scrap tires in resilient playground surfacing. Many schools and recreation facilities are reconfiguring playgrounds to make them more accessible, including the use of resilient surfaces which can be traversed by wheelchairs.

The goal of most scrap tire utilization projects is to find markets for scrap tires so that they do not end up in landfills or on stockpiles. Ironically, one potentially significant use of tires is in the construction and management of landfills. Both shredded and whole scrap tires have been approved in various states for use in constructing leachate beds in landfills. Approval has also been given in some states for the use of shredded tire material as a partial replacement for required daily cover (42).

Scrap whole tires have been used to form artificial fishing reefs, oyster beds, and as a floating breakwater. Goodyear has installed more than 2000 fishing reefs made from old tires. In Ft. Lauderdale, Florida, one of the largest reefs is made of 3×10^3 tires; it stretches ca 2.4 km (53). Tires are also used in playgrounds, flower planters, and shoe soles. The tire splitting industry cuts tires into pieces for gaskets, shims, dock bumpers, shock absorbers, blasting mats, and other articles. Rubber is sliced from tire tread to make strips for floor mats. Approximately 3×10^3 of the estimated 200×10^3 scrap tires generated each year are used as reefs or for splitting.

Tire Retreading and Other Uses

In 1995, about $28\text{--}30 \times 10^6$ tires were retreaded (54). Retreading has been the most cost-effective alternative to recycling rubber. However, worn retreaded tires usually are discarded in a landfill. Approximately 10% of discarded automobile tires and 55–70% of truck and bus tires are retreaded (55).

Economic Aspects

Tire disposal costs are \$0.10–3.00 per tire. Cost for incineration without heat recovery is \$0.35–0.70 per tire. Transportation of discarded tires can cost \$0.04 /kg, and size reduction can cost \$0.20–0.60 /kg. Distribution of passenger car tires is landfill, 85%; retreaded, 10%; and reclaimed, burned for fuel, and split, 5%.

The best use for rubber recycling on a large scale is to provide fuel. Where there is an ample supply of old tires and fees are collected for tire disposal, the economics for using tires for fuel are favorable. Several projects are underway that use tires for fuel, providing process energy and steam for electricity generation (see PROCESS ENERGY CONSERVATION). Cost comparisons are given in Table 6.

Table 6. Fuel Costs^a

Source	Price	Energy, \$ /GJ ^b
coal	\$36.60 /t ³	1.32
No. 6 fuel oil	\$0.079 /L ³	1.89
natural gas	\$153.60 /10 ³ m ²	4.15
rubber		
whole tire	\$0.06 /kg	2.03
ground	\$0.20 /kg	6.77

^a Ref. 54.

^b To convert \$ /GJ to \$ /10 Btu, multiply by 1.054.

BIBLIOGRAPHY

"Rubber" under "Recycling" in *ECT* 3rd ed., Vol. 19, pp. 1002–1010, by J. Paul, *Pedco Environmental, Inc.*

- G. Crane, R. A. Elefritz, E. L. Kay, and J. R. Layman, *Rubber Chem. Technol.* **51**, 577 (1978).
- J. R. Serumgard, personal communication, Scrap Tire Management Council, Washington, D.C., July 6, 1996.
- A-1 Refuse Service, Crow and Sons Sanitary Landfill and Pit, Estes' Service Co., and West Side Sanitary Landfill, Fort Worth, Tex., personal communications, Sept. 3, 1986.
- F. T. Ryan, *Scrap Tires: Alternatives and Markets in the United States*, Goodyear Tire and Rubber Co., Washington, D.C., 1994.
- City officials in Forth Worth, Arlington, and Dallas, Tex., personal communication, Sept. 3, 1986; L. L. Gaines and A. M. Wolsky, *Energy Conservation Through Alternative Uses*, ANLICNSV-5, Argonne National Laboratory, Chicago, Ill., Dec. 1979.
- W. J. Markiewicz and M. J. Granksy, *Solid Waste Management*, Series SW-22, PB203619, U.S. Dept. of Health, Education, and Welfare, Washington, D.C., 1972.
- J. R. Serumgurd, "A New Future for Old Tires—Recycle," presented at the *Chemical Marketing Research Association*, Oct. 1994.
- E. R. Moats, *Resource Recov. Conserv.* **1**(3), 315 (1976).
- M. Weintraub, A. A. Orning, and C. H. Schwartz, *Experimental Studies of Incineration in a Cylindrical Combustion Chamber*, U.S. Bureau of Mines, RI 6908, Dept. of the Interior, Washington, D.C., 1967.
- D. Kennedy, personal communication, Dyna Electronics Corp., McLean, Va., Dec. 1981 and Jan. 1982.
- R. Niles, personal communication, Uniroyal, Inc., Oxford, Conn., June 1981.
- F. Query, paper presented at *National Tire Disposal Symposium*, Washington, D.C., June 1977.
- R. H. Taggart, *Shredded Tires as Auxiliary Fuel*, General Motors, Detroit, Mich., Mar. 1975; W. Underwood, personal communication, General Motors, Detroit, Mich., Jan. 1982.
- C. C. Humpstone, E. Ayres, S. G. Keahey, and T. Schell, *Internat. Res. Technol.*, PB234602 (1974).

15. *Initial Public Offering Perspective*, The Oxford Energy Co., New York, Aug. 19, 1986.

16. M. B. Sikora, *Tire Recovery and Disposal, A National Problem With New Solutions*, Resource Recovery Report, Washington, D.C., June 1986.

17. *A Unique and Effective Waste Tire Treatment Process*, Japan External Trade Organization, Tokyo, Jan. 1980, pp. 32–35.

18. M. Bungay, personal communication, Granular Systems, Sacramento, Calif., Nov. 1986.

19. N. Emanuel, personal communication, Emanuel Tire Co., Baltimore, Md., Oct. 1986.

20. A. T. Kearny, *Scrap Tire Use and Disposal*, Oct. 1992.

21. B. L. Schulman and P. A. White, "Pyrolysis of Scrap Tires Using the Tosco-II Process—Progress Report," ACS Symposium Series No. 76, American Chemical Society, Washington, D.C., 1978.

22. C. E. Haberman, paper presented at *Rubber Division ACS Symposium*, Chicago, Ill., May 1977.

23. L. J. Ricci, *Chem. Eng.* **83**(16), 52 (1976).

24. P. White, personal communication, Tosco Co., Los Angeles, Calif., Dec. 21, 1981.

25. R. Fletcher and H. T. Wilson, *Resource Recov. Conserv.* **5**(4), 333 (1981).

26. J. E. Lunde, personal communication, Intennco, Houston, Tex., July 22, 1981.

27. W. Kaminsky, *Resource Recov. Conserv.* **5**(3), 205 (1980); W. Kaminsky, W. Menzel, and H. Sinn, *Conserv. Recycl.* **1**(1), 91 (1976); W. Kaminsky and H. Sinn, *Kunststoffe* **68**(5), 14 (1978); H. Sinn, W. Kaminsky, and J. Janning, *Angew. Chem. Int. Ed. Engl.* **15**, 660 (1976).

28. L. H. Lewandowski, *Rubber Chem. Technol.* **67**, 447–480 (1994).

29. M. Matsuo, personal communication, Nippon Zeon of America, New York, Dec. 21, 1981.

30. Y. Saeki and G. Suzuki, *Rubber Age* **108**, 33 (Feb. 1976).

31. G. P. Bracker, *Conserv. Recycl.* **4**(3), 161 (1981).

32. P. Bridges, personal communication, Conrad Industries, Centralia, Wash., Nov. 1986.

33. J. A. Beckman, G. Crane, R. A. Elefritz, E. L. Kay, and J. R. Laman, paper presented at the *National Tire Disposal Symposium*, June 14–15, 1977, Washington, D.C.

34. J. W. Laresen and B. Chang, *Rubber Chem. Technol.* **49**, 1120 (1976).

35. G. Crane and E. L. Kay, paper presented at *Rubber Division, ACS Symposium*, Philadelphia, Pa., Oct. 1974.

36. G. Crane, J. W. Fieldhouse, and E. L. Kay, *Rubber Chem. Technol.* **48**, 62 (1975).

37. G. Crane, E. L. Kay, and L. B. Wakefield, *J. Elastomers Plast.* **7**, 372 (1975).

38. G. Crane and E. L. Kay, *Rubber Chem. Technol.* **48**, 50 (1975).

39. Collins, Ramaswamy, Ahmed, and Blumenthal, in Inyang and Bergeson, eds., *Utilization of Waste Materials in Civil Engineering Construction*.

40. *Industrial Recovered Materials Utilization Targets for the Rubber Industry*, report prepared for U.S. DOE, Assistant Secretary for Conservation and Solar Energy, Office of Industrial Programs, Hittman Associates, Inc., 1980.

41. *Data Collection and Analysis Pertinent to the PA's Development of Guidelines for Procurement of Highway Construction Products Containing Recovered Materials*, EPA Contract 68-01-6014, Draft, Vol. 1, Issues and Technical Summary, Franklin Associates, Ltd., and Valley Forge Laboratory, Inc., July 6, 1981.

42. J. R. Sarumgard, *Ground Rubber and Civil Engineering Markets for Scrap Tires*, Aug. 1994.

43. W. O. Murtland, *Elastomerics* **110**, 26 (Mar. 1978).

44. L. J. Ricci, *Chem. Eng.* **84**(14), 71 (1977).

45. D. Tuchman and S. L. Rosen, *Chem. Week* (May 23, 1979).

46. B. D. Bauman, *Rubber Plant News*, **23**, 35–36 (July 4, 1994).

47. J. P. Paul, *CHEMTECH* **9**, 104 (Feb. 1979); D. S. le Beau, *Rubber Chem. Technol.* **40**, 217 (1967).

48. Ref. 45, p. 35.

49. M. Phillips, *Tire Rev.* **94**, 26 (June 1994).

50. J. Miller, *Rubber Plast. News* **23**, 20 (Apr. 11, 1994).

51. R. Woods, *Waste Age*, 54–56, 58, 60, 62, 64 (Aug. 1994).

52. M. Phillips, *Tire Rev.* **94**, 26 (June 1994).

53. *Rubber World*, 67 (Oct. 1978).

54. M. Blumenthal, personal communication, Scrap Tire Management Council, Washington, D.C., Apr. 1996.

55. F. T. Ryan, "Tire Manufacturers' Perspectives on the Legislation Options for Scrap Tire Disposal," paper presented at *Disposal Techniques with Energy Recovery for Scrapped Vehicle Tires Workshop*, Denver, Colo., Feb. 12, 1987.

General References

References 4, 7, 8, 16, 21, 23, 42, and 46 are good general references.

W. J. Search and T. E. Ctvrtnicek, *Resource Recov. Conserv.* **2**(2), 159 (1976).

E. R. Moats, paper presented at the *National Tire Disposal Symposium*, June 14–15, 1977, Washington, D.C.

J. R. Serumgard, Scrap Tire Management Council, Washington, D.C.

F. T. Ryan, State and Regulatory Affairs, Governmental Relations, The Goodyear Tire and Rubber Co., Washington, D.C.

E. F. Sverdrup and B. R. Wendrow, in N. M. Bikales, ed., *The Encyclopedia of Polymer Science and Technology*, Vol. 14, Interscience Publishers, New York, 1989, pp. 787–864.

J. C. Long, *Rubber Recycling, Bibliography, 1989 to September 1994*, ACS Rubber Division Library, Akron, Ohio.

Scrap Tire Users Directory, Recycling Research, Inc.

W. O. Murtiand, *Elastomerics* **109**(12), 39 (1977).

D. J. Zolin, N. B. Fable, and J. F. Gentilcore, paper presented at the *112th Meeting of ACS Rubber Division*, Cleveland, Ohio, Oct. 4–7, 1977.

M. C. Kazarnowicz, E. C. Osmundson, J. F. Boyle, and R. W. Savage, *Chem. Week* (May 23, 1979).

Current Costs for Natural Gas, Coal and Oil, U.S. DOE, Energy Information Administration, National Energy Information Center, Washington, D.C., Jan. 1996.

John P. Paul
Carter & Burgess, Inc.

RECYCLING, WATER.

See WATER, REUSE.

REFRACTION.

See ANALYTICAL METHODS; SPECTROSCOPY, OPTICAL.

REFRACTORIES

Refractories are materials that resist the action of hot environments by containing heat energy and hot or molten materials (1). There is no well-established line of demarcation between those materials that are and those that are not refractory. The ability to withstand temperatures above 1100°C without softening has, however, been cited as a practical requirement of industrial refractory materials (see CERAMICS). The type of refractories used in any particular application depends on the critical requirements of the process. For example, processes that demand resistance to gaseous or liquid corrosion require low permeability, high physical strength, and abrasion resistance. Conditions that demand low thermal conductivity may require entirely different refractories. Combinations of several refractories are generally employed.

Physical Forms

Refractories may be preformed, ie, shaped, or formed and installed on-site, ie, specialties. Innovations in placement and vessel construction has led to a greater emphasis on specialty refractory products. Castables, gunning mixes, and plastic and ramming mixes are used either for repair or for complete new construction of what is known as monolithic linings. The tendency to use monolithics instead of constructions using shaped products has been steadily increasing. As of the mid-1990s, monolithic installations are as common as conventional shaped product construction.

Brick. The standard dimensions of a refractory brick are 229 mm (9 in.) length, 114 mm (4.5 in.) width, and 64 mm (2.5 in.) thickness. This is known as a standard straight. Quantities of bricks are given in brick equivalents, that is, the number of standard straight bricks that have a volume equal to that of the particular installation. The actual shape and size of bricks depends on the design of the vessel or structure in question and may vary considerably from the standard straight brick. For example, bricks for basic oxygen furnaces (BOF), also called BOP or LD vessels, may be in the shape of a key 660 mm long, 76 mm thick, and tapering in width from 152 to 102 mm. Numerous other shapes are available from manufacturers as standard items as well as custom-made or special ordered shapes.

Bricks may be extruded or dry-pressed on mechanical or hydraulic presses. Formed shapes may be burned before use, or in the case of pitch or resin chemically bonded bricks, may be cured (tempered) at a low temperature.

Setter Tile and Kiln Furniture. These products are formed in a similar manner to bricks and are used to support ware during firing operation. The wide variety of available shapes and sizes include flat plates, posts, saggars, and car-top blocks.

Fusion-Cast Shapes. Refractory compositions are arc-melted and cast into shapes, eg, glass-tank flux blocks as large as 305 × 610 × 1219 mm. After casting and annealing, the blocks are accurately diamond ground to ensure a precise fit.

Cast and Hand-Molded Refractories. Large shapes such as burner blocks and flux blocks, and intricate shapes such as glass feeder parts saggars are produced by casting slips, hydraulic cement bonded mixtures, or hand-molding clay or chemically bonded materials. Because these techniques are labor intensive, they are reserved for articles that cannot be satisfactorily formed in any other way, owing to complexity or small production quantities.

Insulating Refractories. Insulating refractories in the form of brick are much lighter than conventional bricks of the same composition by virtue of the brick porosity. The porosity may be introduced by means of lightweight grog or additives that create porosity by a foaming action or by evolution of combustion or decomposition products during the burning process. Refractory fibers (qv) made from molten oxides may be formed into bulk fiber, blankets, boards, or blocks. Such fiber products are used as back-up thermal insulation and low heat-capacity linings in kilns and reheat furnaces (see INSULATION, THERMAL).

Castables and Gunning Mixes. Castables consist of refractory grains to which binders are added. Upon mixing with water, the bond reacts and binds the mass together. Usually this binder is a hydraulic refractory cement. Low (5%) cement, ultralow (<3%) cement, and no-cement castables have been introduced. These castables usually contain chemical deflocculants, dispersants, retarders, and very fine batch ingredients such as silica fume or fine alumina. The products usually contain low water additions. Castables may be placed by vibration, pumping, and in some cases can be made to self-flow (self-leveling castables). The impetus to reduce the cement content in castables has been the desire to improve performance by reducing the low melting eutectics in the lime–alumina–iron–silica system. This has been a principal thrust in refractory development since the mid-1980s.

Gunning mixes are designed to be conveyed pneumatically through a nozzle under water and air pressure. The mixture may be slurried before being shot through the gun or mixed with water at the nozzle. In addition to refractory grains, the mix may contain clay and nonclay additives to promote adherence to the furnace wall. Gunning mixes may be made from cement-bonded refractories, usually based on alumina or aluminosilicate materials, or from basic raw materials which are bonded with silicate or phosphate bonds. The advantage of gunning results from the ability to form large sections quickly without the need for forms. In the case of basic gunning mixes, repairs can be made without removing the furnace from service. That is, the gunning repair can be carried out on hot vessel walls.

Plastic Refractories and Ramming Mixes. Plastic refractories are mixtures of refractory grains and plastic clays or plasticizers with water. Ramming mixes may or may not contain clay and are generally used with forms or steel work. The amount of water used in these products varies but is held to a minimum to facilitate drying. Plastics are generally supplied in a ready-to-use state.

Mortars. Mortars consist of finely ground refractory grains and plasticizers so that a thin coating on the brick can be obtained by dipping. The mortar must be chemically compatible with the refractory lining. This compatibility is established by principles of phase equilibria. The mortars may be designed to bond with the refractories by high temperature reactions during sintering or chemical bonds resulting from additions of sodium silicate, phosphates, cement, or colloidal particles. The conditions needed to cure these chemical bonds range from exposure to air to low temperature heat treatments. Sometimes the heat treatment required to set the bond occurs during the normal furnace or vessel operation, obviating the need for a separate curing procedure.

Composite Refractories. Although many refractories may be considered composites, examples of composites analogous to metal and organic composites are not common (see COMPOSITE MATERIALS). Shaped and specialty refractories containing metallic reinforcements have appeared on the market. Additives such as flake graphite and active metals such as silicon, aluminum, magnesium, and alloys, have been used for nonwetting characteristics and as antioxidants for carbon-containing refractories, respectively.

Refractory Coatings. Refractory coatings (qv) may be applied in a variety of ways. Painting or spraying a fine-grained refractory mix at room temperature yields a dense sintered coating upon heating. Other techniques include thermal or plasma spraying, and chemical vapor deposition. In the first, the powdered coating material is fed into a burner and sprayed at elevated temperatures. The pyroplastic grains form a dense monolithic coating when they impinge on the substrate. Plasma spraying is carried out in essentially the same manner, except that an electrically ionized gas plasma heats the coating powder to temperatures up to 16,700°C.

Raw Materials

In the past, refractory raw materials were used essentially as mined minerals (2–7). Selective mining yielded materials of the desired properties and only in cases of expensive raw materials, such as magnesite, was a beneficiation process required. As of this writing (ca 1995), high purity natural raw materials are increasingly in demand as is synthetically prepared refractory grain made from combinations of high purity and beneficiated raw materials (Tables 1 and 2). The material produced upon firing raw as-mined minerals or synthetic blends is called grain, clinker, co-clinker, or grog. Recycled materials produced by the manufacturers and recovered from users are also being increasingly used to reduce waste. Use of recycled materials has been found to improve refractory performance in some cases when proper care over handling (reclamation) has been exercised. Use of recycled materials is limited in the United States. It is a bit more common in Europe and is expected to increase (see RECYCLING).

Table 1. Composition of Refractory Raw Materials, %^{a, b}

Name	Location	SiO ₂	Al ₂ O ₃	Fe ₂ O ₃	TiO ₂	CaO	MgO	Alkalies
Silica raw materials								
ganister	Bwlehwyn (N. Wales, U.K.)	97.4	0.73	0.78	0.1			
gravel	Sharon Conglomerate, Ohio	98.0	0.3	0.5	0.1			
Clays								
flint clays	Pennsylvania	50.40	34.58	1.42	2.06	0.45	1.00	1.58
	Missouri	43.80	38.29	0.60	2.33	0.07	0.20	0.39
	Kentucky	44.94	35.17	1.56	3.12	0.11	0.18	1.45
	People's Republic of China	51.20	46.62	0.87	0.87			0.15
plastic clays	Pennsylvania	54.00	27.94	1.39	2.45	0.07	1.22	3.33
	Missouri	55.12	28.65	1.66	1.55	0.11	0.10	2.54
	Kentucky	56.42	26.92	2.05	1.95	0.18	0.51	1.75
kaolin	Georgia	43.00	37.55	0.85	2.10	0.09	0.18	0.15
	Florida	46.5	37.62	0.51	0.36	0.25	0.16	0.42
fireclay	Stourbridge (Scotland)	68.1	27.2	1.95	1.1	0.72	0.35	1.28
	Pfalz (Germany)	45.1	36.3	2.21	1.12	0.08	0.11	2.25
plastic	Chasov-Yar (Russia)	51.6	33.3	0.9	1.37	0.53	0.57	3.28
semikaolin	Suvorov (Russia)	46.1	33.9	2.14	1.52	0.41	0.23	0.44
kaolin	Vladimirovka (Russia)	48.3	36.7	0.83	0.78	0.3	0.33	0.77
High alumina								
natural								
siliceous bauxite ^c								
ca 70 ^o o Al ₂ O ₃	Eufaula, Alabama	25.9	70.1	1.13	2.9	0.05	0.03	0.13
ca 60 ^o o Al ₂ O ₃	Eufaula, Alabama	34.9	60.6	1.26	2.5	0.07	0.12	0.11
	People's Republic of China	32.40	63.50	1.50	2.20			0.20
ca 50 ^o o Al ₂ O ₃	People's Republic of China	43.30	52.80	1.42	1.90			0.19
South American bauxite ^c	Guyana	7.0	87.5	2.00	3.25	trace	trace	trace
Chinese bauxite ^c	People's Republic of China	6.0	87.5	1.50	3.75			0.50
kyanite ^d	Virginia	38.6	59–61	0.2–0.9	0.67	0.03	0.01	0.4
sillimanite ^d	India		59–61					
synthetic								
fused alumina								
black		0.48	97.3	0.15	2.45	0.07	0.11	0.05
gray		0.06	99.5 ±	0.15	0.06	trace	trace	0.07
sintered alumina		0.06	99.5	0.06	trace	trace	trace	0.05
sintered mullite	Georgia	27.90	68.00	1.33	2.61	0.06	0.04	0.06
sintered magnesium aluminate	Japan	0.2	67.90	0.2		0.3	31.9	
fused mullite		22.0	77.7	0.12	0.05	trace	trace	0.35
calcium aluminate cement								
low purity		8.4	42.0	10.7		37.0	1.2	
high purity		0.1	79.0	0.3	trace	18.0	0.1	0.5
Zirconium								
zircon ^e		32.5	0.08	0.05	0.06	trace	trace	trace
baddeleyite ^f	Brazil			3–5				
Basic raw materials								
calcined magnesias								
natural magnesite	Austria	0.3	0.3	5.4		2.7	91.3	
	Greece	1.2	0.3	0.3		2.8	95.4	
	U.S.	1.00	0.36	0.65		3.02	94.97	
	People's Republic of China	0.40	0.13	1.73		1.31	96.36	
	Turkey	1.42	<0.02	<0.02		1.58	96.95	
seawater	Japan	0.37	0.04	0.01		1.23	98.32	
	U.K.	0.70	0.20	0.10		2.30	96.66	
	U.S.	0.80	0.12	0.14		2.22	96.70	
	U.S	0.80	0.15	0.22		2.65	96.12	
	Ireland	0.83	0.18	0.17		1.07	97.75	
brine	Israel	0.14	0.02	0.04		0.77	99.45	
dolomite	U.K.	1.8	1.1	5.9		54.6	36.4	
dolomite, low flux	U.S.	0.7	0.3	0.9		57.7	40.4	
chrome ore ^g	Africa	3.8	14.9	27.1		0.3	9.9	
	Philippines	5.5	31.0	15.5		0.5	16.0	

^a Refs. 2 and 5.
^b Difference between total analysis and 100% loss on ignition, %.
^c Calcined.
^d Raw.
^e Contains a trace of Cr₂O₃ and 66.7% ZrO₂.
^f Contains 70.8% ZrO₂.
^g African chrome ore contains 44.0% Cr₂O₃; Philippine, 31.5%.

Table 2. Physical Properties of Refractory Raw Materials^a

Material	Pyrometer cone equivalent	Main crystalline phases	Specific gravity, g cm ³		Apparent porosity, %
			Bulk	True	
Silica					
ganister		quartz		2.66	1.6
gravel		quartz		2.61	0.3
Clays					
flint clays	32–33				
	34	kaolinite, illite, quartz			
	33	kaolinite, quartz, illite	2.55		
plastic clays	31	kaolinite, quartz, illite			
	30	kaolinite, quartz, illite			
	30	kaolinite, quartz, illite			
kaolin	33–34	kaolinite			
fireclay	33–34	kaolinite			
plastic	32–33	kaolinite, illite			
semikaolin	32–33				
kaolin	32–33				
High alumina					
natural siliceous bauxite ^b					
ca 70% Al ₂ O ₃	38–39	mullite	2.85–2.95	3.1–3.2	4–8
ca 60% Al ₂ O ₃		mullite	2.75–2.85 ^c	2.95–3.05	3–7
ca 50% Al ₂ O ₃			2.65		
South American bauxite ^b		corundum, mullite	3.1	3.6–3.7	15–20
Chinese bauxite ^b			3.20		
kyanite ^d	36–37	kyanite		3.5–3.7	
sillimanite ^d		sillimanite		3.23	
synthetic					
fused alumina					
black	42	α-alumina	3.87	4.01	3.49
gray	42 [†]	α-alumina	3.95	3.98	0.5–1.0
sintered alumina	42 [†]	α-alumina	3.45–3.6	3.65–3.80	5.0
sintered mullite	39		2.85		
sintered magnesium aluminate		spinel, periclase	3.33		
fused mullite	39	mullite	3.1	3.45	0.1
calcium aluminate cement					
low purity		calcium monoaluminate			
high purity	34	α-alumina, calcium monoaluminate			
Zirconium					
zircon		zircon		4.2–4.6	
baddeleyite, ZrO ₂		baddeleyite		5.5–6.5	
Basic raw materials					
calcined magnesias					
natural magnesite		periclase	3.2		
		periclase	3.4		
		magnesite, dolomite, calcite	3.40		
		^e	3.39		
seawater		periclase	3.44		
		periclase	3.35		
		periclase	3.40		
		periclase	3.22		
brine		periclase	3.41		
dolomite ^f		periclase + CaO			
chrome ore		periclase + CaO			
		chromite spinel	4.2		
		chromite spinel	3.9		

^a Refs. 2 and 5.
^b Calcined.
^c Another type of ca 60% Al₂O₃ natural siliceous bauxite has a bulk density of 2.70 g cm³.
^d Raw.

^e Unclassified.
^f Both regular and low flux dolomite.

Silica. The most common refractory raw materials are ganister, which is a dense quartzite, and silica gravels (see SILICA; SILICON COMPOUNDS). The latter are generally purer than the former and are often further beneficiated by washing to give the raw material for superduty silica brick having no more than 0.5° impurities (alkalies, Al₂O₃, and TiO₂). Quartzite and gravel deposits are widespread throughout the world. The most important U.S. deposits are found in Pennsylvania, Ohio, Wisconsin, Alabama, Colorado, and Illinois. Synthetically produced electrofused silica is a thermally stable, shock-resistant, high purity raw material.

Fireclay. Fireclays consist mainly of the mineral kaolinite [1318-74-7], Al₂O₃·2SiO₂·2H₂O, with small amounts of other clay minerals, quartzite, iron oxide, titania, and alkali impurities. Clays (qv) can be used in the raw state or after being calcined. Raw clays may be coarsely sized or finely ground for incorporation in a refractory mix. Some high purity kaolins like those that occur in Georgia are slurried, classified, dried, and air-floated to achieve a consistent high quality. The classified clays (qv) also may be blended and extruded or pelletized, and then calcined to produce burned synthetic kaolinitic grog, or coarsely crushed raw kaolinite may be burned to produce grog. Upon calcination or burning, kaolinite decomposes to mullite and a siliceous glass incorporating mineral impurities associated with the clay deposit, eg, quartzite, iron oxide, titania, and alkalies, and is consolidated into dense hard granular grog at high temperatures. Fireclay deposits are widely distributed; in the United States they occur in Pennsylvania, Missouri, and Kentucky.

Bauxitic Kaolins and Mullites. Deposits of bauxitic kaolins, kaolins having aluminous minerals, have been discovered that have alumina contents between 50 and 70° . These materials are made into refractory aggregates called calcines, grog, clinker, or grain. In addition to selectively mined deposits, synthetic compositions can be prepared from kaolin and alumina and other minerals to produce compositions of desired alumina and mineralogical content. These synthetic mullites are readily available in the form of sintered and fused aggregates.

High Alumina. The naturally occurring raw materials are bauxites, sillimanite [12141-45-6/ group minerals, and diaspore clays (see ALUMINUM COMPOUNDS). Other high alumina raw materials are made by beneficiation, blending, and other processing techniques.

Bauxites. Bauxite [1318-16-7] consists mainly of gibbsite [14762-49-3], Al(OH)₃, and varying amounts of kaolinite, and iron and titania impurities. Because the loss on ignition is high, bauxite must be calcined to high temperatures before use. During calcination, it is converted to a dense grain consisting mainly of corundum [12252-63-0], Al₂O₃, and mullite. Refractory-grade bauxite is relatively rare because a high iron content makes most bauxites unsuitable for refractory use. Commercially mined deposits are in South America, especially Guyana and Surinam, and the People's Republic of China. Other deposits occur in India and Central Africa but are not mined for refractory grades at present.

Sillimanite Minerals. This group includes sillimanite, andalusite [12183-80-1], and kyanite [1302-76-7]. All have the formula Al₂O₃·SiO₂. Upon heating, a mixture of mullite, silica, and a siliceous glass is obtained. The specific gravity of sillimanite and andalusite is ca 3.2, and that of kyanite is 3.5. Thus, kyanite expands upon conversion to mullite by ca 16–18 vol ° , but sillimanite and andalusite expand only slightly. Kyanite is found in India and South Africa, and in the United States in Virginia and South Carolina. Large-grained kyanite is rare, and the largest size commercially available from U.S. sources is ca 500 µm (35 mesh). Refractory-grade sillimanite occurs in India and South Africa.

Mullite. Although mullite is found in nature, for example, as inclusions in lava deposits on the island of Mull, Scotland, no commercial natural deposits are known. It is made by burning pure sillimanite minerals or sillimanite–alumina mixtures. Fused mullite of high purity is obtained by arc-melting silica sand and calcined alumina. High purity sintered mullite is made from alumina and silica, but requires mineralizing agents and very high temperatures.

Alumina. A pure although not necessarily a refractory grade of alumina is obtained from bauxite by the Bayer process. In this process, the gibbsite from the bauxite is dissolved in a caustic soda solution and thus separated from the impurities. Alumina, calcined, sintered, or fused, is a stable and extremely versatile material used for a variety of heavy industrial, electronic, and technical applications.

Calcined alumina is a reactive powder used to make synthetic grain. It also may be used as a bonding or fine component in batched refractory mixes, as a raw material for molten cast refractories, or for refractory casting slips.

Sintered alumina, also known as tabular alumina, is formed by burning aggregates made from reactive calcined alumina to high temperatures to obtain a stable high purity corundum grain. Fused alumina is obtained by fusing either calcined alumina or bauxite. Bauxite is beneficiated during fusion as iron and silica are removed as ferrosilicon. A special grade of fused alumina is obtained by blowing air through a stream of molten alumina. These hollow spheres of alumina, called bubble alumina, make an excellent medium weight aggregate of high purity and refractoriness (see ALUMINUM COMPOUNDS).

Calcium Aluminate Cements. Low purity calcium aluminate [12042-78-3] cements are obtained by sintering or fusing bauxite and lime in a rotary or shaft kiln. A high purity calcium aluminate cement, 2CaO·5Al₂O₃, capable of withstanding service temperatures of 1750°C can be prepared by the reaction of high purity lime with calcined or hydrated alumina (see ALUMINUM COMPOUNDS).

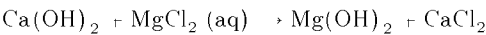
Zirconia. Zircon (zirconium silicate), the most widely occurring zirconium-bearing mineral, is dispersed in various igneous rocks and in zircon sands. The main deposits are in New South Wales, Australia; Travancore, India; and Florida in the United States. Zircon can be used as such in zircon refractories or as a raw material to produce zirconia. The zircon structure becomes unstable after about 1650°C, depending on its purity, and decomposes into ZrO₂ and SiO₂ rather than melting (see ZIRCONIUM AND ZIRCONIUM COMPOUNDS).

Zirconia occurs as the mineral baddeleyite [12036-23-6], for instance, in a deposit around Sao Paulo, Brazil. However, these baddeleyite deposits generally contain large amounts of impurities and only about 80° ZrO₂. High purity zirconia can be obtained from baddeleyite by leaching with concentrated sulfuric acid or chlorination at high temperatures. The zirconium sulfates and chlorides thus formed are readily separated from the impurities. Zirconia also is made from zircon by electric-arc melting under reducing conditions. Here, the silica is separated from the zirconia by adding iron to form ferrosilicon. The most remarkable property of zirconia is its volume instability. Upon heating, unstabilized ZrO₂ transforms from the low temperature monoclinic form to the tetragonal form at about 1000°C, resulting in a 9° volume contraction. By adding impurities such as magnesia, calcia, or yttria, zirconia can be transformed into a cubic crystal phase stable at all temperatures, although often only enough impurities are added for partial stabilization, ie, both cubic and either the tetragonal or monoclinic forms exist.

Basic Raw Materials.

Magnesite. Calcined or dead-burned magnesite [13717-00-5] is obtained by firing naturally occurring magnesium carbonate to 1540–2000°C. This treatment produces a dense product composed primarily of periclase [1309-48-4], MgO. Large sedimentary deposits of magnesite occur in Austria, Manchuria, Greece, the Ural Mountains, and in the United States in Washington, Nevada, and California. Calcia, silica, alumina, and iron-bearing phases occur as accessory minerals (see MAGNESIUM COMPOUNDS).

Synthetic magnesia is most often produced from seawater, known as seawater magnesia. Seawater contains approximately 1294 ppm Mg. Synthetic magnesia can also be produced from brine wells or lakes which have much higher concentrations of magnesium. Regardless of the source of magnesium, the sea or brine water is treated with hydrated lime, Ca(OH)₂, that precipitates Mg(OH)₂:



The Mg(OH)₂ precipitate is filtered, dried, and calcined to produce caustic magnesia. The caustic magnesia is then mechanically compacted and sintered to form a dense granular product. Large seawater or brine plants are located in the United States (Gulf of Mexico, California, Michigan, and New Jersey), Mexico, the United Kingdom, Ireland, Israel, Italy, Japan, and Russia. Magnesium hydroxide also occurs in sedimentary deposits as the mineral brucite [1317-43-7], eg, in Quebec and Nevada.

Dolomite. Dolomite [17069-72-6], CaMg(CO₃)₂, occurs in widespread deposits in many areas including southern Austria, the U.K., Russia, and the United States. Raw dolomite may be used for certain refractories, but in most instances it is calcined to form a grain consisting primarily of MgO (periclase) and CaO [1305-78-8]. Calcined dolomite absorbs H₂O and CO₂ from the atmosphere and eventually disintegrates. Fluxes such as SiO₂, Fe₂O₃, and Al₂O₃ increase hydration resistance but also sharply reduce the fusion point of the dolomite. High purity, low flux dolomites with less than 2°

impurities are produced by high temperature calcination of natural dolomites in Ohio and Pennsylvania (see LIME AND LIMESTONE).

Forsterite. Pure forsterite is rare in nature. Most natural magnesium orthosilicates form solid solutions of fayalite, Fe₂SiO₄, and forsterite. Forsterite refractories are usually made by calcining magnesium silicate rock such as dunite, serpentine, or olivine with sufficient magnesia added to convert all excess silica to forsterite and all sesquioxides to magnesia spinels.

Chrome Ore. Chromia-bearing spinel materials, although considered neutral, are generally used in combination with basic magnesite. Chrome ores consist essentially of a complex solid-solution series of spinels including hercynite, FeO·Al₂O₃; ferrous chromite, FeO·Cr₂O₃; magnesioferrite, MgO·Fe₂O₃; picrochromite, MgO·Cr₂O₃; spinel, MgO·Al₂O₃; and magnetite, FeO·Fe₂O₃. Silicate phases such as serpentine, talc (qv), and enstatite are commonly associated with the spinel grains. The principal deposits of chrome ore occur in Africa (Transvaal, South Africa, Zimbabwe), Russia, Turkey, Greece, Cuba, and the Philippines; African and Russian chrome ores are high in iron content, and Cuban and Philippine ores are higher in alumina.

Silicon Carbide. Silicon carbide is made by the electrofusion of silica sand and carbon. Silicon carbide is hard, abrasion resistant, and has a high thermal conductivity. It is relatively stable but has a tendency to oxidize above 1400°C. The silica thus formed affords some protection against further oxidation (see CARBIDES).

Beryllia and Thoria. These are specialty oxides for highly specialized applications that require electrical resistance and high thermal conductivity. Beryllia is highly toxic and must be used with care. Both are very expensive and are used only in small quantities.

Carbon and Graphite. Carbon (qv) and graphite [7782-42-5] have been used alone to make refractory products for the lower blast furnace linings, and electrodes for steel and aluminum production. They are also commonly used in conjunction with other refractory raw materials. These materials are highly refractory nonwetttable materials and are useful refractories in nonoxidizing environments. Carbon blacks are commercially manufactured, whereas graphite for refractory use has to be mined.

General Properties

Oxides. A number of simple and mixed refractory oxide materials are described in Table 3. The most widely used simple oxide is Al₂O₃. It has moderate thermal shock resistance, good stability over a wide variety of atmospheres, and is a good electric insulator at high temperatures. The strength of ceramics is influenced by minor impurities and microstructural features. In general, polycrystalline alumina has reasonably good and nearly constant strength up to about 1000–1100°C; at higher temperatures, the strength drops to much less than one-half of the room temperature value over a 400°C temperature increment. Single-crystal alumina is stronger than polycrystalline Al₂O₃ and actually increases in strength between 1000 and 1100°C. Fused silica glass has excellent thermal shock properties but devitrifies on long heating above 1100°C and loses much of its shock resistance.

Table 3. Properties of Pure Refractory Materials^a

Material	CAS Registry Number	Formula	Mp, °C	True specific gravity, g cm ³	Mean specific heat		Thermal conductivity, W (m·K)		Linear thermal expansion coefficient per°C × 10 ⁶ , from 20–1000°C
					J (kg·K) ^b	Temp range, °C	500°	1000°	
							C	C	
aluminum oxide	[1344-28-1]	Al ₂ O ₃	2015	3.97	795.5	25–1800	10.9	6.2	8.6
beryllium oxide	[1304-56-9]	BeO	2550	3.01	1004.8	25–1200	65.4	20.3	9.1
calcium oxide	[1305-78-8]	CaO	2600	3.32	753.6	25–1800	8.0	7.8	13.0
magnesium oxide	[1309-48-4]	MgO	2800	3.58	921.1	25–2100	13.9	7.0	14.2
silicon dioxide ^c	[7631-86-9]	SiO ₂		2.20	753.6	25–2000	1.6	2.1	0.5
thorium oxide	[1314-20-1]	ThO ₂	3300	10.01	251.2	25–1800	5.1	3.0	9.4
titanium oxide	[13463-67-7]	TiO ₂	1840	4.24	711.8	25–1800	3.8	3.3	8.0
uranium oxide	[1344-58-7]	UO ₂	2878	10.90	251.2	25–1500	5.1	3.4	
zirconium oxide ^d	[1314-23-4]	ZrO ₂	2677	5.90	460.6	25–1100	2.1	2.3	6.5–10
mullite	[55964-99-3]	3Al ₂ O ₃ ·2SiO ₂	1850°	3.16	628.0	25–1500	4.4	4.0	4.5
spinel	[1302-67-6]	MgO·Al ₂ O ₃	2135	3.58	795.5		9.1	5.8	6.8
forsterite	[15118-03-3]	2MgO·SiO ₂	1885	3.22	837.4		3.1	2.4	9.5
zircon	[10101-52-7]	ZrO ₂ ·SiO ₂	2340–2550 ^f	4.60	544.3		4.3	4.1	4.0
carbon	[7440-44-0]	C		2.10	1046.7	25–1300	13.4	9.9	4.0
silicon carbide	[409-21-2]	SiC	3990 ^g	3.21	795.5	25–1300	22.5	23.7	5.2

^a Refs. 8–10.

^b To convert J to cal, divide by 4.184.

^c Silica glass.

^d Cubic, stabilized with CaO.

^e Congruent.

^f Incongruent.

^g Dissociates above 2450°C in reducing atmosphere and is readily oxidized above 1650°C.

Beryllium and magnesium oxides are stable to very high temperatures in oxidizing environments. Above 1700°C MgO is highly volatile under reducing conditions and in vacuum, whereas BeO exhibits better resistance to volatilization but is readily volatilized by water vapor above 1650°C. Beryllia has good electrical insulating properties and high thermal conductivities; however, its high toxicity restricts its use. Calcium oxide, and to a lesser extent uranium oxide, hydrate readily. The latter, UO₂, can be oxidized to lower melting U₃O₈. Zirconia in pure form is rarely used in ceramic bodies; however, stabilized or partially stabilized cubic ZrO₂ is the most useful simple oxide for operations above 1900°C. Thorium oxide exhibits good properties for high temperature operation but is expensive and radioactive. Titanium oxide is readily reduced to lower oxides and cannot be used in neutral or reducing atmospheres.

Carbon, Carbides, and Nitrides. Carbon (graphite) is a good thermal and electrical conductor. It is not easily wetted by chemical action, which is an important consideration for corrosion resistance. As an important structural material at high temperature, pyrolytic graphite has shown a strength of 280 MPa (40,600 psi). It tends to oxidize at high temperatures, but can be used up to 2760°C for short periods in neutral or reducing conditions. The use of new composite materials made of carbon fibers is expected, especially in the field of aerospace structure. When heated under

oxidizing conditions, silicon carbide and silicon nitride [12033-89-5], Si₃N₄, form protective layers of SiO₂ and can be used up to ca 1700°C (see NITRIDES).

Silicon carbide has very high thermal conductivity and can withstand thermal shock cycling without damage. It also is an electrical conductor and is used for electrical heating elements. Other carbides have relatively poor oxidation resistance. Under neutral or reducing conditions, several carbides have potential usefulness as technical ceramics in aerospace application, eg, the carbides (qv) of B, Nb, Hf, Ta, Zr, Ti, V, Mo, and Cr. Ba, Be, Ca, and Sr carbides are hydrolyzed by water vapor.

Silicon nitride has good strength retention at high temperature and is the most oxidation resistant nitride. Boron nitride [10043-11-5] has excellent thermal shock resistance and is in many ways similar to graphite, except that it is not an electrical conductor.

Borides and Silicides. These materials do not show good resistance to oxidation. Some silicides, however, form SiO₂ coatings upon heating which retards further oxidation. Molybdenum disilicide [1317-33-5], MoSi₂, is used widely, primarily as an electrical heating element.

Metals. The highest melting refractory metals are tungsten (3400°C), tantalum (2995°C), and molybdenum (2620°C). All show poor resistance to oxidation at high temperatures. Hafnium–tantalum alloys form a tightly adhering oxide layer that gives partial protection up to 2200°C. The layer continues to grow upon extended use.

Phase Equilibria. Phase diagrams represent the chemical equilibria that exist among one, two, or three components of a system under the influence of temperature and pressure (11,12). Reference to a phase diagram permits the determination of the amount and composition of solid and liquid phases that coexist under certain specified conditions of temperature and pressure for a particular system. Using such information, the occurrence of physical and chemical changes within a system or between systems at high temperatures can be predicted. Systems containing more than three components are difficult or impossible to present graphically; however, mathematical methods have been suggested (13,14).

Phase diagrams can be used to predict the reactions between refractories and various solid, liquid, and gaseous reactants. These diagrams are derived from phase equilibria of relatively simple pure compounds. Real systems, however, are highly complex and may contain a large number of minor impurities that significantly affect equilibria. Moreover, equilibrium between the reacting phases in real refractory systems may not be reached in actual service conditions. In fact, the successful performance of a refractory may rely on the existence of nonequilibrium conditions, eg, environment (15–19).

Physical Properties. The important physical properties of some refractories are listed in Tables 4–7. Brick bulk density depends on the specific gravity of the constituents and the porosity. The latter is controlled by the porosity of the raw materials and the brick texture. Coarse, medium, and fine-sized material contribute to the degree of particle packing. Usually the highest density possible is desired. Upon firing, the grains and matrix form glassy, direct, or solid-state ceramic bonds. Sintering is generally accompanied by shrinkage, unless new components are formed that may cause expansion. Differential volume change between the coarse and fine fraction caused by differential sintering rates and formation of addition phases may create stresses. Particle size distribution, forming method, and firing process contribute to texture, whereas permeability is related to porosity, which in turn is dependent upon texture.

Table 4. Composition of Alumina, Silica, and Zirconia Refractories, %^a

Type	SiO ₂	Al ₂ O ₃	Fe ₂ O ₃	TiO ₂	CaO	MgO	Alkalies
silica	95–96	0.2–1.5	0.8–1.0	0.3–0.1	3.3–2.7	<0.1	<0.2
fireclay							
semisilica	70–79	17–27	0.6–2.0	0.8–1.6	0.1–0.4	0.1–0.4	0.2–0.4
medium duty	56–70	25–36	1.8–3.4	1.3–2.7	0.2–0.4	0.4–0.6	1.0–2.7
high duty	51–59	35–41	1.5–2.5	1.1–3.0	0.2–0.6	0.1–0.6	1.3–2.6
super duty	50–54	40–46	0.8–2.3	1.1–2.5	0.1–0.5	0.4–0.6	0.1–1.4
high alumina, Al ₂ O ₃							
50° o	40–47	47–57	0.9–1.6	2.2–2.4	0.5–0.6	0.3–0.6	0.8–1.3
60° o	27–37	58–67	0.9–2.7	1.7–3.0	0.1–0.3	0.2–0.7	0.2–1.2
70° o	19–28	68–77	0.9–2.2	2.0–3.3	0.1–0.3	0.1–0.7	0.2–1.2
80° o	8–17	78–87	0.7–1.7	2.5–3.2	0.1–0.2	0.1–0.4	0.1–1.0
85° o ^b	6–10	82–86	1.0–2.5	2.0–3.0	trace–0.2	trace–0.4	0.1–0.3
90° o	3–10	88–96	0.1–1.4	0.1–2.6	0.1–0.2	0.1–0.3	trace–0.9
100° o	0.2–1.1	97–99	0.1–0.3	trace–0.3	0.1–0.3	0.1–0.3	0.1–0.3
zircon ^c	32	1	0.1	1.8	trace	0.1	0.02
molten cast							
Al ₂ O ₃ –ZrO ₂ –SiO ₂ ^d	11–13	50–51	trace–1.0	trace–1.0	trace	trace	1.0–1.5
Al ₂ O ₃ –SiO ₂	18–22	57–70	1–4	3.0–4.5	0–1	0–1	0.0–1.5
Al ₂ O ₃ high soda	0.04	93.8	0.1	0.05	0.1	0.1	5.6
Al ₂ O ₃ low soda	0.10	99.2	0.1	0.05	0.1	0.1	0.3

^a Refs. 2, 5, and 20–23.
^b Phosphate bonded.
^c Contains 64° o ZrO₂.
^d Contains 33–35° o ZrO₂.

Table 5. Physical Properties of Alumina, Silica, and Zirconia Refractory Brick^a

Type	PCE ^b	Modulus of rupture, MPa ^c	Deformation under load ^d		Linear reheat change ^e		Thermal spalling loss ^f		Bulk density g cm ³	Porosity, ° o
			Temp, °C	Change, ° o	at °C	° o	at °C	° o		
silica		2.8–11.2	1650	0					1.60–1.80	20–30
fireclay										
semisilica	27–31	2.1–4.2	1450	0.1–2			1480	0–10	1.80–2.10	20–30
medium duty	29–31	7.0–11.2	1450	1–6			1600	2–6	2.11–2.20	17–21
high duty	31–33	2.8–21.0	1450	0.5–15	1600	0–1 S	1600	3–7	2.13–2.30	4–30
super duty	33–34	2.8–24.0	1450	0–9	1600	0–1 S	1650	2–8	2.28–2.48	5–22
high alumina, Al ₂ O ₃										

50 ^o °	34–35	7.0–11.2	1450	2–6	1600	0–1 E	1650	8–12	2.27–2.43	14–18
60 ^o °	36–37	7.0–11.2	1450	1–4	1600	0–5 E	1650	1–5	2.10–2.49	13–28
70 ^o °	37–38	7.0–11.2	1450	1–3	1600	1–7 E	1650	3–7	2.20–2.66	14–28
80 ^o °	38–39	7.0–12.6							2.50–2.90	14–29
85 ^o ° ^g	38–39	21.0–35.0	1450	0.3–3	1600	0–2 E	1650	0–5	2.70–2.92	12–17
90 ^o °	40–41	14.0–35.0	1700	0–1	1700	0–1 E	1650	0–1	2.67–3.10	11–27
100 ^o °	41–42	12.6–21.0	1650	1–2	1700	0–0.5 S	1650	0–5	2.84–3.10	19–29
zircon			1600	2–5	1540	0			3.77	20
molten cast										
Al ₂ O ₃ -ZrO ₂ -SiO ₂									3.70–3.74	0.8–3
Al ₂ O ₃ -SiO ₂									3.00–3.20	1–3
Al ₂ O ₃ high soda									2.89	
Al ₂ O ₃ low soda									3.50	

^a Refs. 2, 5, and 20–23.

^d Load 172 kPa; ^o ° of linear change after 1.5 h.

^e S shrinkage; E expansion.

^f As determined by ASTM C38.

^b PCE pyrometer cone equivalent, as determined by ASTM C24.

^c To convert MPa to psi, multiply by 145; for kPa, multiply by 0.145.

^g Phosphate bonded.

Table 6. Composition of Basic Refractory Bricks, %^a

Type	SiO ₂	Al ₂ O ₃	Fe ₂ O ₃	CaO	MgO	Cr ₂ O ₃	Residual carbon
magnesite							
burned	0.4–4.5	0.1–1.0	0.1–2.3	0.6–3.8	91.0–98.0	0.0–0.9	
burned, tar impregnated	0.4–4.5	0.1–1.0	0.1–2.3	0.6–3.8	91.0–98.0	0.0–0.9	2.0–2.5
tar bonded, tempered	0.5–3.8	0.2–1.0	0.1–3.0	0.9–6.2	88.0–98.0	0.0–0.4	4.2–4.7
resin bonded	0.5–3.8	0.2–1.0	0.1–0.5	1.5–2.7	96.0–97.0	0.0–0.2	3.8–4.7
high carbon	0.7–2.3	0.1–0.7	0.2–0.9	0.9–3.0	94.0–98.0		7.0–20.0
magnesite–chrome							
burned	1.8–7.0	6.0–13.0	2.0–12.0	0.6–1.5	50.0–82.0	6.0–15.0	
direct bonded	1.0–2.6	3.0–16.0	3.0–10.0	0.6–1.1	50.0–80.0	7.0–20.0	
chemical bonded	2.0–8.0	5.0–14.0	2.0–8.0	1.0–2.0	50.0–80.0	6.0–15.0	
chrome							
burned	5.0–8.0	27.0–29.0	12.0–20.0	0.4–1.0	15.0–23.0	29.0–35.0	
chrome–magnesite							
burned	3.0–5.0	8.0–20.0	8.0–12.0	0.7–1.1	40.0–50.0	18.0–24.0	
dolomite							
burned	0.8–1.5	0.3–0.8	0.6–1.5	38.0–58.0	38.0–58.0		0.0–1.4
tar bonded	0.3–1.5	0.1–0.6	0.2–2.0	50.0–58.0	38.0–43.0	0.0–0.3	4.0–5.0

^a Refs. 2, 5, 7, and 20–24.

Table 7. Physical Properties of Basic Brick^a

Type	Bulk density, g cm ³	Apparent porosity, ^o °	Modulus of rupture, MPa ^b		Refractoriness under load, shear temp, °C	Linear reheat change ^c at 1650°C, ^o °
			20°C	1260°C		
magnesite						
burned	2.8–3.0	15–19	7.0–24.5	3.5–18.5	1590–1760	0–4 S
burned, tar impregnated ^d	3.0–3.2	(13–17) ^e	20.0–35.0	10.5–21.0		
tar bonded, tempered	3.0–3.1	3–7	7.0–11.0		1700 ±	
resin bonded	2.9–3.1	4–7	8.0–27.0			
high carbon	2.7–3.0	1–6	7.0–9.0			
magnesite–chrome						
burned	2.9–3.0	17–20	3.0–4.9	0.7–2.0	1500–1700	0–0.3 S
direct bonded ^f	2.9–3.2	14–19	5.6–13.9	8.4–17.5	1700 ±	0.7 S–1.0 E
chemical bonded	3.0–3.2	18–20	7.0–14.0	0.7–2.8	1650–1760	1 S–5.0 S
chrome						
burned	3.1–3.3	18–20	6.0–14.0	0.4–1.1	1260–1370	0–0.6 S
chrome–magnesite						
burned	3.0–3.2	19–21	5.6–8.4	2.8–11.2	1650–1700	0–0.1 E
dolomite						
burned	2.7–3.1	6–19	7.0–32.8		1700 ±	0–0.2 S
tar bonded	2.8–3.0	(8–12) ^e	7.0–11.0			

^a Refs. 25.
^b To convert MPa to psi, multiply by 145.
^c S = shrinkage; E = expansion.
^d 7.0–16.0 MPa at 1480°C.
^e Obtained on ignited sample.
^f 2.0–4.2 MPa at 1480°C.

Mechanical Properties. The physical properties of a particular refractory product depend on its constituents and manner in which these were assembled. The physical properties may be varied to suit specific applications. For example, for thermal insulations highly porous products are employed, whereas dense products are used for slagging or abrasive conditions.

The transverse strength (modulus of rupture) at room temperature is related to the degree of bonding. Fine-grained refractories generally are stronger than coarse-grained types and those having a low porosity are stronger than those of high porosity. However, the room temperature strength of a refractory is not necessarily indicative of the strength at high temperature because the bond strength may be the result of a glassy phase that softens upon heating.

Generally, high temperature strength is lower than room temperature strength. The former is a measure of the degree of solid-state bonding between refractory grains, whereas high temperature creep indicates the amount of associated liquid or glassy phases and their viscosity. The development of solid-state or direct-bonded basic brick requires high firing temperatures, and is impeded by glassy phases. By referring to phase diagrams, refractory compositions may be designed that avoid the development of such phases.

The modulus of elasticity (MOE) is related to the strength and can be used as a nondestructive quality control test on high cost special refractory shapes such as slide gate valves employed in the pouring of steel (qv). The slide gate type must be selected to ensure chemical compatibility and it must be used in a way to reduce thermal shock. The performance of a properly selected and used slide gate is directly related to its strength and therefore predicted by its MOE.

Thermal Properties. Refractories, like most other solids, expand upon heating, but much less than most metals. The degree of expansion depends on the chemical composition. A diagram of the thermal expansion of the most common refractories is shown in Figure 1.

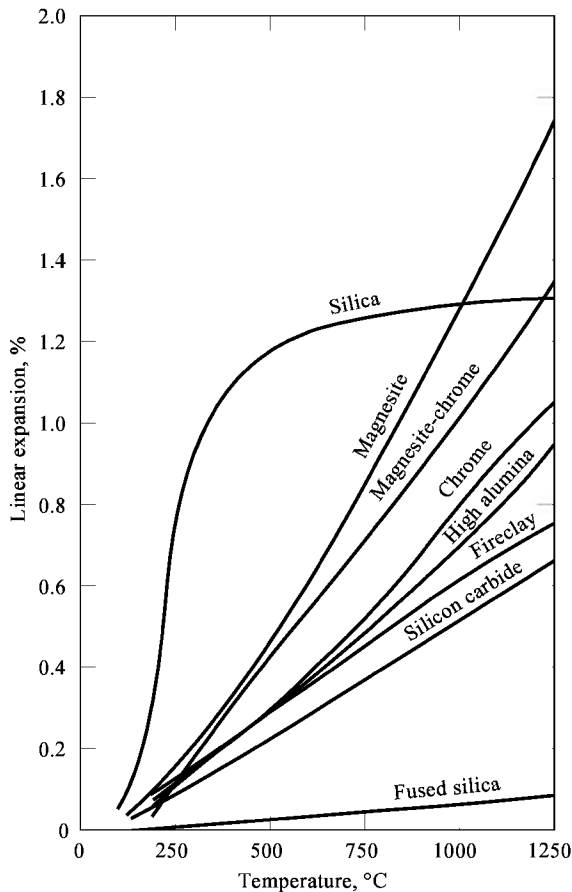


Fig. 1. Thermal expansion values of some materials.

Courtesy of Chemical Engineering (28).

Reheat Change. Most refractory bricks are not chemically in equilibrium before use. During the prolonged heating in service, additional reactions occur that may cause the brick to shrink or expand. For instance, 70% Al_2O_3 bricks consist of a mixture of corundum grains, mullite, and siliceous glass. In service, the siliceous glass reacts with corundum to form additional mullite, causing expansion. Mullite bricks that contain only mullite are volume stable because equilibrium conditions have been reached during initial firing. Considerable expansion also may be caused by gas formation during heating, for instance, by the decomposition of sulfates. In the presence of a viscous siliceous glass, these gases cannot escape and expansion occurs. This mechanism explains the bloating behavior of common bloating ladle brick. The linear reheat changes of various refractories during an American Society for Testing and Materials (ASTM) reheat test are shown in Tables 5 and 7.

In general, basic brick exhibits good volume stability at high temperatures. Some slight shrinkage may occur in chemically bonded or low fired brick; this shrinkage is most pronounced in high silicate compositions. In high fired, high purity brick containing chrome ore, some expansion on reheat is generated by periclase–spinel reactions, usually with very complex interdiffusion mechanisms, where the spinel constituents are dissolved in the periclase lattice at high temperature and the spinel phases precipitate from the periclase solid solution on cooling. The expansion may continue with repeated heating and cooling cycles as well as oxidation reduction cycles.

Thermal Conductivity. The refractory thermal conductivity depends on the chemical and mineral composition of the material and increases with decreasing porosity. The thermal conductivities of some common refractories are shown in Figure 2.

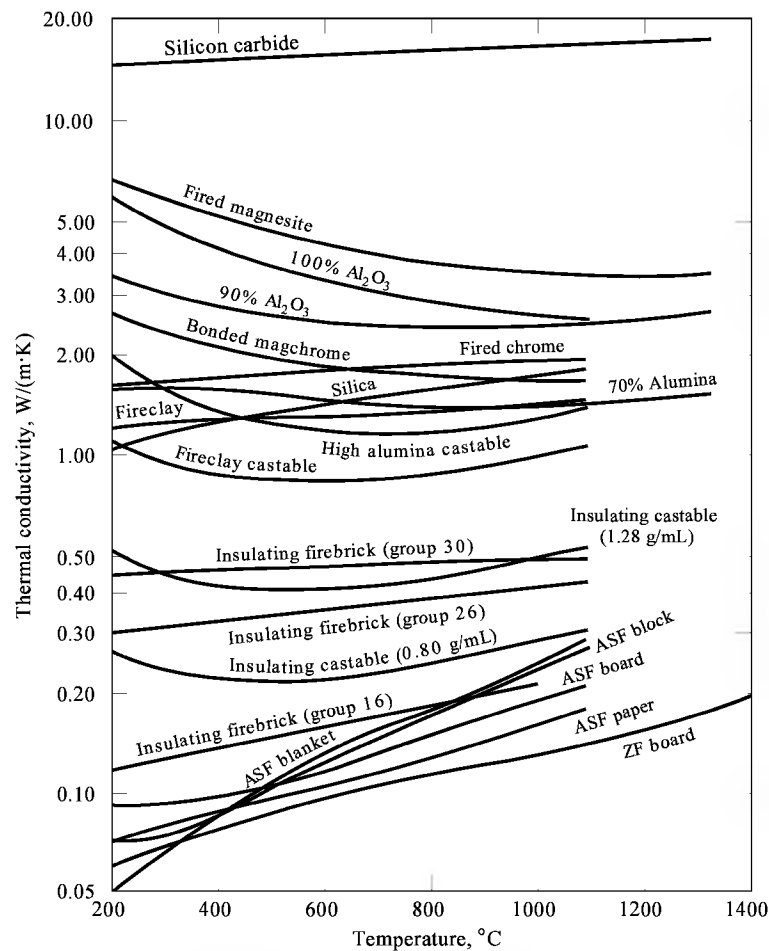


Fig. 2. Thermal conductivity of refractories where ASF = aluminosilicate fiber and ZF = zirconia fiber. See Table 13 for group classifications (5,25).

Specific Heat. In some applications refractories are used for heat-exchange purposes on the regenerative principle, for instance, in blast-furnace stoves. High heat capacity is required in such applications (Table 8).

Table 8. Mean Specific Heats of Refractory Brick and Minerals, Between 0°C and the Indicated Temperature, J (kg·K)^{a,b}

Temp, °C	Brick					Mullite	Cristobalite	Periclase	Corundum
	Fireclay	Silica	Magnesite	Chrome	Forsterite				
0	249.5	218.5	268.9	219.8	232.9	237.9	213.3	268.9	221.1
93	257.3	243.0	283.1	227.5	258.5	248.2	236.6	293.5	253.4
204	266.4	272.8	300.0	235.3	279.3	276.7	263.8	310.3	276.7
316	274.1	296.1	312.9	243.1	297.4	288.3	309.0	324.5	292.2
427	284.5	307.7	324.5	250.8	310.3	296.1	324.5	332.3	303.9
538	293.5	318.1	333.6	257.3	318.1	301.3	333.6	338.8	312.9
649	302.6	325.8	340.1	259.9	323.3	306.4	341.4	345.2	320.7
760	311.6	333.6	346.5	268.9	328.4	310.3	346.5	349.1	325.8
871	320.7	336.2	353.0	274.1	333.6	312.9	350.4	354.3	332.3
982	327.1	341.4	359.5	279.3	338.8	316.8	353.0	358.2	336.2
1093	333.6	346.5	365.9	284.5	343.9	319.4	355.6	362.0	341.5
1204	338.8	351.7	372.4	287.0	349.1	322.0	358.2	364.6	345.2
1316	343.9	356.9	378.8	289.6	354.3	324.5	359.5	368.5	349.1
1427	347.8	360.7	384.0	292.2	359.5	327.1	360.7	372.4	353.0
1538						329.7	362.0	375.0	356.9
1642						331.0	363.3	377.6	360.7
1760						333.6		380.1	364.6

^a Ref. 26.
^b To convert J (kg·K) to Btu (lb·°F), multiply by 2.39 × 10⁻⁴.

Thermal Spalling. Refractories are brittle and stresses caused by sudden variations in temperatures can cause cracking and destruction. The susceptibility to thermal cracking and spalling depends on certain characteristics of the raw material and the macrostructure of the particular refractory. Spalling resistance may be increased by either preventing cracks from forming or preventing cracks from growing. The approach used to effect spall resistance determines which properties of the refractory are optimized. Generally low thermal expansion, high density and high thermal conductivity, and high strength refractories exhibit good thermal shock resistance. Fireclay and high alumina refractories usually have a higher resistance to thermal shock than periclase refractories. Dense strong bodies withstand high stress and transmit it over large volumes; when failure occurs it is serious. Weak porous bodies, however, tend to crack before catastrophically large stresses are generated and thus are much less seriously damaged and generally remain intact.

The resistance against thermal spalling of fireclay and high alumina brick is indicated in Table 5. No standard test has been adopted for basic brick. Refractories composed of 100% magnesia exhibit poor thermal shock resistance, which is improved by addition of chrome ore. So-called direct bonded basic brick, composed of magnesia and chrome additions, exhibits good thermal shock resistance.

Refractoriness. Most refractories are mixtures of different oxides, sometimes with significant quantities of impurities. Thus, they do not have sharp melting points but a softening range. Refractoriness is the resistance to physical deformation under the influence of temperature. It is determined by the pyrometric cone equivalent (PCE) test for aluminosilicates and resistance to creep or shear at high temperature (see ANALYTICAL METHODS).

Manufacture

Processing. Initial processing may include an extensive survey of the deposit, selective mining, stockpiling by grade, and beneficiation techniques such as weathering, grinding, washing, heavy-media separation, froth flotation, etc. Some materials can be used without further processing although many must be subjected to heat treatment. In the case of synthetic grain, the selected and beneficiated raw materials are blended in the desired proportions and formed into suitable shapes for calcination by briquetting, pelletizing, or extrusion. Slurries may be calcined; however, this practice is avoided in the interest of fuel economy. Originally, calcination referred specifically to the treatment of calcareous minerals to remove CO₂. The term has come to be used to indicate heat treatment to sinter or burn (dead burn) the refractory grain to a stable dense material as well as to decompose minerals. Calcination may be carried out to rotary kilns, shaft kilns, multiple-hearth furnaces, or fluidized-bed reactors. The last two devices are reserved for relatively light calcining. The particular feed is dictated by the kiln type and the precalcination processes to which the raw materials or mixes may have been subjected. These hard, burned materials are called grain, clinker, or grog, a term also used for ground firebrick. Low density or expanded aggregates can be made by burning clay or clay mixtures that evolve gas during burning and thereby expand or produce porosity. The burn-out material may be naturally present in the clay or mixed with the clay before burning.

Both raw and processed materials can be fused or melted in electric-arc furnaces. These materials can be melted and then cast into shapes, formed in the furnace itself as an ingot, or formed into fibers. Partially fused ingots are crushed into grains. Because these ingots are not homogeneous, the grains have to be graded; the middle portion is purer than the outside material which is less well fused. Melts used for casting or fiber forming are homogeneous.

Crushing and Grinding. Some raw materials, such as hard clay and quartzite, must first be crushed to grains small enough for the grinding equipment. In general, a jaw, gyrotory, or roll crusher is employed (see [SIZE REDUCTION](#)).

Almost all raw materials require grinding after primary crushing. For coarse grinding, a dry pan or occasionally a wet pan is used. The dry pan is similar to a grist mill but has a perforated bottom through which the crushed material is continuously removed. The wet pan is similar, but has a solid bottom. For very fine grinding, a ring-roll, ball, or impact mill is employed.

Screening. To obtain a high density product, the mix is made from materials that have been sized into classes by means of standard screens. In a continuous screening operation, the ground raw materials are generally fed to vibrating high capacity screens that may be heated. Material that does not pass the screen is returned to the grinding system for further size reduction. Coarse and medium fine-grain sizing is accomplished by the aforementioned methods, whereas fine-sized materials generated in rod mills, ball mills, ring-roll mills, etc, are classified by air separators. A single screen with uniform openings produces a single product referred to as a straight grind. For closer control of a mix, screen analysis-sized fractions are used to produce the overall desired particle size distribution. Here a series of two, three, or more vibrating screens are set in stacks yielding a coarse band, an intermediate band, and a fine band. These fractions are blended into the mix in the proper proportions to provide the desired particle packing. The distribution may be continuous following either distributions proposed by Furnas, Andreason, or the modified Andreason distribution or may be discontinuous or semicontinuous gap-grain sizing. The use of a particular distribution depends on a variety of factors such as the degree of plasticity of the mix constituents, the method of forming to be used, and the requirement for density and thermal shock resistance. A typical brick formulation may contain 50% coarse (ranging from 0.625 cm to fine sand) and about 20% less than 44 μm (similar to fine flour) material.

Mixing. As in other ceramic processes, more than one type of raw material is often required for a refractory product. The purpose of mixing is to uniformly distribute the various ingredients (see [MIXING AND BLENDING](#)). Although the specific steps and equipment involved in the mixing of batches for fireclay, high alumina, and basic refractories are somewhat different, the general principles are similar. Mixes that are to be dry-pressed contain 2–6% binding liquid, depending on the plasticity of the raw material bond system and the fineness of the mix. The ingredients may be blended in a pug mill, dry pan, or other type mixer and tempered with the bonding ingredients. Tempering, in the sense used here, denotes the kneading action produced on the mix, usually in a muller mixer. Mixes to be extruded or hand formed contain 10–20% liquid. These mixes may be prepared in a pug mill or wet pan.

The mixing of tar-, resin-, and chemically bonded brick material presents special problems. Tar- and resin-bonded mixes are usually basic compositions. Because coal-tar and petroleum pitches have ring-and-ball values of ca 100°C, ie, the softening point as determined by ASTM D30, provisions for mixing the ingredients at elevated temperatures must be made. Typically, the grain is heated to ca 150°C and maintained at that temperature while the hot pitch is added. After a preset mixing time, the batch is transferred to the pressing equipment. For resin-bonded mixes, the grain does not have to be heated, although liquid resins may be heated to enhance their flow characteristics. Phenolic (phenol–formaldehyde), alkyd–oil–urethane, urea–formaldehyde, furan resins, etc, are used (see [ALKYD RESINS](#); [AMINO RESINS AND PLASTICS](#); [PHENOLIC RESINS](#)). Resin-bonding systems are more expensive than pitches, but concerns regarding potential health risks of pitch fumes have renewed interest in resins.

Chemically bonded basic bricks are blended much the same as burned brick mixes except that a bonding agent, eg, magnesium sulfate or magnesium chloride, is added to the mix as well as tempering water to form oxysulfates or oxychlorides.

Forming. Most refractory shapes are formed by mechanical equipment, but some very large or intricate shapes require hand molding in wooden, steel-lined molds with loose liners to permit easy removal of plaster of Paris molds.

A few fireclay refractories are produced by the stiff-mud process using an auger machine that pugs, de-airs, and continuously extrudes a clay column. A wire cutter cuts the clay into blanks which are then sized, shaped, and branded by a repress machine.

Refractory shapes are generally produced on a mechanical toggle press, screw press, or hydraulic press. In this operation, a mold cavity is filled with the damp mix. During the pressing cycle, some products such as fireclay brick are de-aired by applying a vacuum to the top and bottom press heads which contain a series of small openings. De-airing promotes a denser product and reduces lamination. In some basic presses, bricks are encased in steel or have internal steel plates which are placed in the mold box before charging; the plates become an integral part of the brick upon pressing. Plates may also be bonded to the brick after pressing. From one to four bricks may be pressed at a time, depending on the size. Pressures range from ca 17 (2500) to ca 98 MPa (14,000 psi) for plastic firebrick or nonplastic basic mixes, respectively. Pitch-bonded brick must be formed hot. Initial heating of the press may be required, but the residual heat from the mix is usually sufficient to maintain the necessary temperature. Another type is the impact- or jolt-mold press that consists of a mold box and press heads activated by pneumatic cylinders, eg, jack hammers. Some special shapes are produced by air-ramming which is similar to hand molding, except that reinforced steel molds are required and the damp mix is slowly fed to the mold while molders manually compact the material with pneumatic rammers. Special shapes can also be formed by slip casting and hot pressing.

Fusion-cast refractories are formed by first melting the material in an electric arc. The liquid is poured into a mold and allowed to cool. Various annealing processes and refining techniques ensure a uniform and structurally sound shape. The case form is then cut or ground to size.

Fused-cast refractory is very dense but may contain a system of closed pores and large, highly oriented grains may exist in a particular casting. The size and distribution of the pore and grain phases must be controlled.

Isostatic pressing gives a highly uniform product, although the production rate is somewhat low. It typically contains very small grains and little or no porosity. In this process, a rubber sock or bag of the desired shape is filled with the refractory mix. The sock is then subjected to extremely high pressure in a hydraulic pressure chamber.

Large and small shapes may be slip cast from both plastic and nonplastic mixes by the usual techniques. Precise shapes, such as glass feeder parts, are made in this way as well as large flux blocks. The process requires the formulation of a slip of suitably stable character to be poured into a plaster mold to be dewatered. After it solidifies, the mold is removed and dried further before firing.

Drying. The drying (qv) step for large shapes is critical. Extremely large fireclay and silica shapes are sometimes allowed to dry on a temperature-controlled floor heated by steam or air ducts embedded in the concrete. Smaller shapes are generally dried in a tunnel dryer. The ware is placed on cars that enter the cold end and exit at the hot end. These dryers may be humidity-controlled and are heated by their own source or with waste heat from the burning operation. Microwave and infrared drying are being investigated.

Pitch-bonded and resin-bonded bricks are treated or cured in special ovens at temperatures higher than that used for drying other types of bricks. Pitch-bonded brick is cured at 230–320°C; this process is called tempering which is not to be confused with tempering during mixing. Tempering removes some of the volatiles from the pitch and eliminates or reduces thermoplastic or slumping behavior in service (see [DRYING](#)).

Curing. Some chemically bonded bricks require some elevated heat treatment that is typically higher than the tempering process mentioned above, but less temperature than that required to form ceramic bonds. One example is aluminosilicate brick bonded with phosphoric acid. A very strong

bond can be developed with heat treatment temperatures above 600°C. Even though adequate strengths can be obtained at low temperatures, complete curing of even these refractories may be desired for certain applications.

Burning. Bricks are fired or burned in kilns to develop a ceramic bond within the refractory and attain certain desired properties. This step does not apply to chemically or organically bonded products. Preferred for this purpose is the continuous or tunnel kiln, a structure of narrow cross section and 61–183 m in length. The bricks are set on cars that move slowly through the kiln, which is divided into the preheating, burning, and cooling zones. The length of the zones controls the rates of heating and cooling, and the time at temperature (soak time). The temperature in the firing zone ranges from 1000 to 1700°C. Gas, oil, or coal may be used as a fuel. Where the quantities of refractories being fired is too small for use of a tunnel kiln, periodic or shuttle kilns are used.

Silica brick and large fireclay shapes are fired in circular downdraft kilns. These kilns vary in diameter and can accommodate up to 150,000 23-cm bricks or their equivalent in other sizes. The complete burning cycle for a typical periodic kiln ranges from 21 to 27 days as compared with four to seven days for a tunnel kiln.

The shuttle kiln consists of a firing chamber with two or more kiln cars on which the bricks to be fired are set. While one load of brick is being fired, a second is being set. Somewhat similar is the bell top or top-hat kiln which is raised and lowered above and over the kiln cars to be fired. These kilns are more expensive to operate than tunnel kilns but provide flexibility in burning conditions and production schedules.

Burned brick may be impregnated with tar or pitch to improve corrosion resistance. The heated bricks are placed in the impregnation unit which is sealed and evacuated to remove air from the pores of the refractory. The evacuated chamber is then filled with hot pitch and the vacuum is released. The treated product is allowed to drain free of excess pitch and is ready for shipment. Although this treatment is primarily used on basic refractories, it can be extended to other classes. The benefits derived from impregnation are lost if the pitch is burned off at high temperature; therefore, a reducing atmosphere is required such as is encountered in a basic oxygen steelmaking furnace.

Specialty Refractories. Bulk refractory products include gunning, ramming, or plastic mixes, granular materials, and hydraulic setting castables and mortars. These products are generally made from the same raw materials as their brick counterparts.

Granular materials are shipped raw or calcined and usually have been ground to a specified screen size or size distribution. The additives depend on the application and service conditions. These materials are used in construction, repair, or maintenance of furnaces and vessels. Refractory mortars are used to lay brick of the same composition. These are manufactured wet premixed or dry.

Ramming mixes and plastics are manufactured similarly to brick; that is, the coarse refractory aggregate is added to a wet pan and combined with a small amount of raw clay or organic agents for plasticity. Water is added to the batch; refractory plastics generally contain more water and raw clay than ramming mixes. After discharge from the wet pan, plastic mixes are formed by extrusion into 25-kg blocks and sliced into rectangular slabs. The material was historically packaged damp and ready for installation with pneumatic rammers. Dry plastic mixes requiring on-site mixing are available. Ramming mixes are generally shredded into small lumps and placed damp in airtight steel drums ready for installation with pneumatic rammers. Ramming mixes are also available in dry form for mixing on-site. When installed, ramming mixes generally require forms to contain the material as it is rammed. Plastics, however, are soft and cohesive and can be installed without forms, although supporting anchors for vertical and in some cases horizontal installations are used.

Basic raw materials are susceptible to hydration and therefore specialty products are shipped dry and mixed with water on-site for gunning or ramming. Certain basic specialties are offered with organic vehicles such as oils and can be used without on-site mixing. Information on manufacturing can be found in References 26–30.

Economic Aspects

The principal consumers of refractories are the iron (qv) and steel (qv) industries (Table 9). The decrease in refractories consumption coincides with technological changes in the manufacture of steel. First, in the 1960s, open-hearth steelmaking was replaced by the basic oxygen furnace (BOF). Steady improvements in BOF practice and improvements in refractory composition and design has led to improved refractory performance. Early BOF furnaces would be campaigned for 100 heats before relining. As of the mid-1990s, 3000 heats and more are typical. More sophisticated ladle metallurgical practice has been employed leading to improved steel quality and improved refractory performance. The values of shipped refractories from 1970 through 1993 are given in Table 10.

Table 9. Distribution of U.S. Refractory Sales^a

Industry	U.S. sales, % of total				
	1964	1977	1979	1985	1986 ^b
iron and steel	61.2	47.2	51.6	49.4	49.0
nonferrous metals	5.4	6.1	7.5	8.2	7.5
cement	2.0	3.6	4.9	4.3	4.1
glass	5.3	4.6	5.1	4.7	6.4
ceramics	5.0	8.7	9.7	8.1	8.3
chemical and petroleum	3.0	2.5	2.1	3.3	3.5
public utilities	1.1	0.7	0.9	1.2	0.8
export	5.3	6.5	7.4	4.9	5.1
all other	11.7	20.1	10.9	16.4	15.2
Total, \$ × 10 ⁶	497.0	1299.0	1727.0	1580.0	1520.0

^a Ref. 31.

^b Last year production data reported by industry; see ASTM C28.

Table 10. Refractories Shipped, 1970–1993 10⁶^a

Year	Refractory		Total
	Clay	Nonclay	
1970	256	342	599
1973	327	453	780
1976	464	621	1084
1979	712	1015	1727
1980	512	698	1211
1985	695	888	1582
1990	772	1232	2202
1993	749	1163	1912

^a Ref. 31.

The number and dollar sales values of brick equivalents of shaped refractories are given in Table 11. There was a large drop in refractory shipments at the beginning of the 1980s that corresponded to a reduction in steel plant capacity by almost half during that period. Large steelmaking complexes were called integrated shops because they were designed to integrate the ironmaking and steelmaking of a steel product stock in one large complex. Since the 1980s these shops have given way to more specialized, much smaller steel shops called minimills. Steel production has increased from the 1980 level but the main change has been the quality improvement of steel resulting from the change to ladle metallurgical processes such as ladle refining, vacuum processing, and the growth of the so-called minimills, based on electric furnace shops instead of the large integrated steel plants (see STEEL).

Table 11. Shipments of Shaped Refractories^{a, b}

Year	Fireclay ^c		High alumina ^d		Silica		Basic ^e	
	10 E	10 \$	10 E	10 \$	10 E	10 \$	10 E	10 \$
1965	614	102	56	41	110	21	140	124
1975	507	142	129	137	41	32	126	279
1985	172	167	98	234	6	15	54	247
1993	144	164	118	231	21	32	76	418

^a Ref. 31.
^b E = brick equivalents.
^c Includes regular fireclay, semisilica superduty fireclay, ladle brick, and insulating firebrick (IFB) below 23.
^d Includes high alumina, mullite, and extra high alumina brick.
^e Includes magnesite, magnesite–chrome, chrome, chrome–magnesite brick, and dolomitic brick.

ASTM Classifications and Specifications

Classifications. In addition to testing methods, ASTM publishes a list of classifications covering a wide variety of refractory types (32). The various brands from numerous producers and producing districts are grouped into classes using a nomenclature indicative of chemical composition, heat resistance, and service properties. The number of refractories being developed has been large. The classifications may be inadequate to encompass all refractories encountered, but these represent the only universally accepted standards specifying composition or properties. A more general classification of refractory form and type is also available (33).

Fireclay and High Alumina Brick. ASTM designation C27 covers fireclay and high alumina brick (Table 12). High alumina brick is classified according to alumina content, starting at 50% and continuing up to 99% Al₂O₃. Manufacturers are allowed 2.5% of the nominal alumina content, except for 85 and 90% (2.0%) and 99% (min 97%). An additional requirement for alumina bricks with Al₂O₃ content of 50, 60, 70, and 80%, are pyrometer cone equivalents (PCEs) of 34, 35, 36, and 37, respectively.

Table 12. ASTM C27 Fireclay Brick Classification^a

Classification	Type	PCE ^b	Panel spalling loss, max, %	Cold modulus of rupture, MPa ^c	Other test requirements
superduty	regular ^d	33	8 at 1650°C	4.14	bulk density, min, 2.243 g cm ³
	spall-resistant ^d	33	4 at 1650°C	4.14	
	slag-resistant	33		6.89	
high duty	regular	31½	10 at 1600°C		bulk density, min, 2.195 g cm ³ , or max porosity 15%
	spall-resistant	31½		3.45	
	slag-resistant	31½		8.27	
semisilica ^e				2.07	silica content, min, 72%
medium duty		29		3.45	
low duty		15		4.14	

^a Ref. 34.
^b PCE = pyrometer cone equivalent.
^c To convert MPa to psi, multiply by 145.
^d Reheat shrinkage at 1600°C is 1.0% max.
^e Hot-load subsidence at 1350°C is 1.5% max.

Basic Brick. Chrome brick, chrome–magnesite brick, magnesite–chrome brick, and magnesite brick are classified under ASTM C455. There are six classes of chrome–magnesite and magnesite–chrome brick starting with 30% MgO and increasing in 10% increments to 80% MgO. The minimum requirement for MgO is 5% less than the nominal percentage MgO specified for each class. For the three magnesite classes for 90, 95, and 98% MgO the minimum requirement is 86, 91, and 96% MgO, respectively. A chrome brick is manufactured entirely of chrome ore.

Insulating Brick. ASTM classifies insulating firebrick under C155 by group. The group number corresponds to the service temperature divided by 100 (Table 13). For example, group 16 corresponds to a test temperature of ca 1600°F (871°C).

Table 13. ASTM C155 Insulating Brick Classification^a

Group identification number	Test temperature for <2% reheat change, °C (°F)	Bulk density, < g cm ³
16	815 (1550)	0.545
20	1065 (1950)	0.641
23	1230 (2250)	0.768
26	1400 (2550)	0.865

28	1510 (2750)	0.961
30	1620 (2950)	1.089
32	1730 (3150)	1.522
33	1790 (3250)	1.522

^a Ref. 34.

Mullite Refractories. Mullite refractories are classified under ASTM C467. This brick must have an Al₂O₃ content between 56 and 79° and contain less than 5° impurities. Impurities are considered metal oxides other than those of aluminum and silicon. The hot-load subsidence is 5° max is 1593°C.

Silica Brick. Under ASTM C416, types A and B silica bricks are classified according to chemical composition and strength. Silica brick must have an average modulus of rupture of 3.5 MPa (500 psi), <1.5% Al₂O₃, no more than 0.2° TiO₂, <2.50% Fe₂O₃, and <4.00% CaO. Type A brick must have a flux factor <0.5. The flux factor is equal to the percent alumina plus twice the percent of alkalis. Type B are all other silica brick covered by the standard chemical and strength specifications.

Zircon Refractories. ASTM C545 classifies zircon refractories in two types. Types A and B have the same chemical requirements of not less than 60° ZrO₂ and not less than 30° SiO₂. Type A (regular) must have a density of less than 3.85 g cm² and type B (dense) more than 3.85 g cm².

Castable Refractories. Hydraulic-setting refractory castables are classified under ASTM C401 into five classifications: Regular Castable Refractory, Low Cement Castable Refractories, Ultra-Low Cement Castable Refractories, No-Cement Castable Refractories, and Insulating Castable Refractories. The regular castable class contains hydraulic setting cement and has a total calcium oxide content greater than 2.5°. This class is subdivided into a normal strength and high strength class which have strengths, as measured by modulus of rupture, of 2.07 and 4.14 MPa (300 and 600 psi), respectively. Each is further classified into subgroups A through G on the basis of shrinkage as shown (34).

Class	5-h Firing temp after which <5% shrinkage permitted, °C
A	1095
B	1260
C	1370
D	1480
E	1595
F	1705
G	1760

Low Cement, Ultra-Low Cement, and No-Cement Castables are classified on the basis of calcium oxide content. These are 1–2.5, 0.2–1.0, and 0.2° CaO maximum, respectively. In the latter case the lime content is not a result of a hydraulic setting cement constituent but comes from aggregate impurities. The insulating class is also subdivided. This division is shown in Table 14. Refractories used in steel-pouring pits are classified under ASTM C435 (Table 15).

Table 14. Classification of Insulating Castables ^a

Class	5-h Firing temp after which <5% shrinkage permitted, °C	Density, g cm ³
N	925	0.88
O	1040	1.04
P	1150	1.20
Q	1260	1.44
R	1370	1.52
S	1480	1.52
T	1595	1.60
U	1650	1.68
V	1760	1.68

^a Ref. 34.

Table 15. ASTM C435 Classification For Steel-Pouring Pit Refractories ^a

Class	Type	PCE ^b	Porosity ^c , °	Reheat change ^c above 1350°C, ° ^d
nozzle	A	15–20	8	1.0
	B	20–29	8	1.0
	C	29 ^c	10	1.0 ^e
sleeve	A	15–20	10	1.0
	B	20–29	10	1.0
	C	29 ^c	10	1.0 ^e
laddle brick ^f	A	15 ^c	18 ^e	5.0 ^g
	B	15 ^c	18 ^e	2.5
	C	26 ^c	18 ^e	0.5 ^{e,h}

^a Ref. 34.

^b PCE pyrometer cone equivalent.

^c Value is minimum unless otherwise indicated.

^d Except ladle brick which is heated as stated; min refers to expansion, and max to shrinkage of diameter.

^e Value is maximum.

^f Modulus of rupture of 4.83 MPa (700 psi) required.

^g Above 1290°C.

^h Above 1500°C.

Specifications. Among the many specifications covering refractory products, the best known are those published by ASTM. In addition, specifications are issued by the U.S. Government and the armed forces. The former are generally preceded by the prefix HH and the later by the prefix MIL. The ASTM refractory specifications always suggest a use, whereas federal and military specifications are inconsistent in this respect.

No less important are specifications issued by industrial consumers. Consumers are making purchasing decisions on the basis of the manufacturer's processing capability and requiring the inspection of production control charts and data.

The International Organization for Standardization (ISO) has issued the ISO 9000 series standards for quality management and quality assurance requirements and guidance. These standards are being rapidly adopted as national standards. Companies wishing to compete within regional trading groups are being required to apply these standards to their manufacturing operations (see MATERIALS STANDARDS AND SPECIFICATIONS; QUALITY ASSURANCE).

Analytical and Test Methods

The test methods applicable to refractories are available (34). These methods are summarized in Table 16.

Table 16. ASTM Test Methods for Refractories^a

Material	Test identification	Properties
burned brick	C20, C830	apparent porosity, water adsorption, bulk density
brick, various shapes	C133, C607, C93	crushing strength, modulus of rupture
basic brick	C456	hydration resistance
brick and tile	C154	warpage
granules	C357, C493	bulk density
periclase grains	C544	hydration
mortar	C198	cold-bonding strength
air-setting plastics	C491	modulus of rupture
castables	C298	modulus of rupture
granular dead-burned dolomite	C492	hydration
fireclay plastics	C181	workability index
castables	C417	thermal conductivity
plastics	C438	thermal conductivity
general refractories	C288	disintegration in CO atmosphere
	C135	true specific gravity
	C201	thermal conductivity
	C92	sieve analysis and water content

^a Ref. 34.

Refractoriness. Refractoriness is determined by several methods. The pyrometric cone equivalent (PCE) test (ASTM C24) measures the softening temperature of refractory materials. Inclined trigonal pyramids (cones) are formed from finely ground materials, set on a base, and heated at a specific rate. The time and temperature (heat treatment) required to cause the cone to bend over and touch the base is compared to that for standard cones.

The standard ASTM PCE test is relative and used extensively only for alumina–silica refractories and raw materials (see Table 5). However, the upper service limit is generally several hundred degrees below the nominal PCE temperature because some load is generally applied to the refractory during service. In addition, chemical reactions may occur that alter the composition of the hot face and therefore the softening point. The relationship between PCE numbers and temperature is described in ASTM C24.

Another measure of refractoriness is the hot-compressive strength or hot-load test for refractory bricks or formed specialties. The specimen carries a static load from 69 kPa (10 psi) to 172 kPa (25 psi). It is heated at a specific rate to a specific temperature which is then held for 1.5 h, or it is heated at a specific rate until it fails. The percent deformation or the temperature of failure is measured. The procedure is described in ASTM C16.

Thermal Strength and Stability. Dimensional changes that occur upon reheating can be determined by ASTM C605, C210, C179, or C113. Specimens are selected and cut or formed to an appropriate size and measured before and after being heated at an appropriate temperature schedule to a specified temperature for 5 to 24 hours. The linear, diametral, or volume percentage change is noted.

High temperature strength of refractory materials is determined on rectangular prisms 25 × 25 × 150 mm cut from the product being tested. The specimens are placed in a furnace, heated to a desired temperature, and the modulus of rupture is determined. A detailed description is given in ASTM C583.

Thermal spalling resistance is determined by ASTM C439, C122, C107, and C38. The last three methods apply to fireclay, high alumina bricks, and plastics using the apparatus described in C38. The specimens are weighted and built into a panel (wall section) that was preheated for 24 h at a specific temperature. The panels are then subjected to 12 cycles of heating to 1400°C followed by cooling with water and air spray. The panel is dismantled, the loose spalls are removed, and the weight loss is recorded as percent spalling loss. Thermal spalling of silica brick is determined on six specimens placed on a guarded hot plate. The specimens are heated and cooled at a specified rate. The heating rate is reported and any cracking that may have occurred is described. Thermal spalling tests are designed to determine the resistance to thermal shock. An additional test (ASTM C1100), the Ribbon Thermal Shock Test, measures the elastic modulus and strength of aluminosilicate refractories before and after exposure to a thermal shock.

Special Tests. Even though the American Society for Testing and Materials offers a wide range of test methods, there are other special tests that are imposed upon the manufacturer by consumers, the military, the U.S. Government, and in some cases local or municipal governments. These tests are generally very specific and are oriented toward particular service conditions. In many instances, the producers develop special tests within their laboratories to solve customer problems or predict product or production performance. Many of these tests subsequently are adopted by ASTM.

Health and Safety Factors

Industrial refractories are by their very nature stable materials and usually do not constitute a physiological hazard. This is not so, however, for unusual refractories that might contain heavy metals or radioactive oxides, such as thorium and uranium, or to binders or additives that may be toxic.

Inhalation of certain fine dusts may constitute a health hazard. For example, exposure to silica, asbestos, and beryllium oxide dusts over a period of time results in the potential risk of lung disease. OSHA regulations specify the allowable levels of exposure to ingestible and respirable materials. Material Safety Data Sheets, OSHA form 20, available from manufacturers, provide information about hazards, precautions, and storage pertinent to specific refractory products.

Manufacturers have recognized that products that exhibit a potential impact on the health and safety of users require greater sensitivity of handling. Safe use, disposal, and recycling of these products is the result. An example of such a product stewardship program is that developed by the refractory ceramic fiber (RFC) coalition. This program embraces the entire life cycle of the ceramic fiber products. The aim is to prevent unacceptable risks to those working with the fiber products. Similar programs are being developed in other refractory manufacturing segments.

Selection and Uses

Any manufacturing process requiring refractories depends on proper selection and installation. When selecting refractories, environmental conditions are evaluated first, then the functions to be served, and finally the expected length of service. All factors pertaining to the operation, service design, and construction of equipment must be related to the physical and chemical properties of the various classes of refractories (35).

Service conditions that impair effectiveness of refractories include chemical attack by, eg, slags, fumes, gases, etc; operating conditions, ie, temperatures and cycling; and mechanical forces, ie, abrasion, erosion, and physical impact. Design factors that influence selection include equipment type and construction, ie, brick or monolithic material; refractory function, ie, material containment, flow deflection, heat storage or release; heat environment, ie, exposure to constant or variable temperatures; refractory strength, ie, exposure to varying stress conditions; and thermal function, ie, insulation, dissipation, or transmission of heat.

The effects of processing conditions on refractories may be summarized as follows (35):

Service condition	Chemical resistance
oxidizing atmosphere	oxides and combinations of oxides (ie, silicates, fireclays) are unaffected; carbon and graphite oxidize; silicon carbide is fairly stable to 1650°C
steam or water vapor	can cause hydration of magnesite refractories at low temperatures and oxidizes carbon and graphite above 705°C
hydrogen	silica and silica-containing refractories are attacked above 1100°C; high alumina, ZrO ₂ , MgO, and calcium–aluminate refractories show good resistance
sulfur and sulfates	above 870°C sulfur reacts with refractories containing silica; carbon and high purity oxides show good resistance; sulfates react to some degree; calcium–aluminate cement is more resistant than Portland cement
fuel ash reducing atmosphere	alkali and vanadium attack from ash can be severe on fireclay; high alumina resists most refractories are stable; however, iron oxide impurities, when reduced, can cause destruc-tion, particularly if cycled
carbon monoxide	iron impurities can act as a catalyst to cause deposition of carbon in fireclay refractories; CO can oxidize graphite and SiC and cause destructive changes in basic refractories
chlorine and fluorine	chlorine attacks silicates above 650°C; F attacks all refractory materials except graphite; basic refractories have poor resistance to both
acids	basic refractories have fair to poor resistance, fireclay and high alumina good resistance, except for HF; zircon, zirconia, and silicon carbide have good resistance; carbon and graphite do not react
alkalies	fireclay and high alumina perform well at low temperatures; magnesite refractories, fair to good; chrome refractories, poor; and graphite, excellent

Standardized applications of different refractory types are listed in an excellent series compiled by the ASTM (36). These surveys cover the principal industrial applications of refractories and furnish a description of furnace operations and destructive influences such as slagging, erosion, abrasion, spalling, and load deformation.

By far the most common industrial refractories are those composed of single or mixed oxides of Al, Ca, Cr, Mg, Si, and Zr (see Tables 1, 4, and 6). These oxides exhibit relatively high degrees of stability under both reducing and oxidizing conditions. Carbon, graphite, and silicon carbide have been used both alone and in combination with the oxides. Refractories made from these materials are used in ton-lot quantities, whereas silicides are used in relatively small quantities for specialty application in the nuclear, electronic, and aerospace industries.

The common industrial refractories are classified into acid, SiO₂ and ZrO₂; basic, CaO and MgO; and neutral, Al₂O₃ and Cr₂O₃. Oxides within each group are generally compatible with each other, whereas mixtures of acid and basic oxides often give low melting products. Neutral oxides are generally compatible with both acidic and basic oxides.

Reactions Between Refractories and Liquids. Molten metals are generally much less reactive than slags. Therefore, the response of a refractory to a chemical environment generally depends on its slag resistance which, in turn, depends on the compositions and properties of slag and refractory. Other factors include temperature, severity of thermal cycling or shock of the process, velocity and agitation of the slag in contact with the refractory, and the abrasion to which the refractory is subjected. Considering these factors, it is not surprising that similar refractories placed in similar furnaces can wear at vastly different rates under different operation practices.

The basicity of slags and refractories is determined by the ratio of the basic oxides to acid oxides. Most often this ratio is taken to be the ratio of the lime to silica, ie, the most influential basic and acid oxides, respectively. More complex formulas are commonly used, however. As a general rule, acid slags (CaO / SiO₂ <1) require acid refractories (see Table 4); basic slags (CaO / SiO₂ >1) require basic refractories (see Table 6). Fireclay and aluminosilicate refractories perform best for slags having a CaO / SiO₂ ratio <1; however, for acid slags containing considerable amounts of iron or manganese oxides, high alumina refractories are required. High alumina refractories are also superior to fireclay refractories when the basicity-to-acidity ratio = 1. When the CaO / SiO₂ ratio > 1, basic refractories such as MgO, MgO–CaO, and MgO–Cr₂O₃ should be employed. Magnesium oxide resists slags of a wide range of compositions; however, in actual practice all basic refractories contain some silica which usually occurs in various silicate phases at the grain boundaries.

As slag attack proceeds, liquid phases migrate from the hot face to the cooler regions and attack of the grain-boundary silicate phase precedes solution of the periclase grain. For this reason, both the character of the bonding silicate phase and the porosity of the refractory are important. Penetration of the slag may be impeded by materials such as carbon, pitch, or resins by lowering the porosity and changing the wetting characteristics.

Slags penetrate refractories until the viscosity becomes too high for further migration. Refractories having high thermal conductivity can cause the slag to chill and thus the penetrated volume is reduced. High thermal gradients within a refractory often reduce the effect of slag. Therefore, efforts to increase the thermal conductivity of refractories and create high thermal gradients to refractories exposed to high slag conditions have been made. Use of internal metal plates within refractories have been tried. Incorporating large flake graphite in the magnesium oxide refractories is common. The success of this approach has led to the use of flake graphite in other oxide systems including alumina and spinel. Using water-cooled panels behind these high thermal conductivity refractories to increase the thermal gradient has also been tried. The technique was so successful that water cooling panels without refractories are common in electric furnaces above slag lines.

Slag penetration alters the structure of the refractory by changing its porosity, density, mineralogical makeup, and strength. If the altered refractories are subject to thermal cycling or if volume changes occur upon crystallization of the slag, stress concentrations build up immediately behind the densified zone and spalling or cracking may result. An example of structural alternation is the iron oxide bursting phenomenon in magnesite–chromite brick. Iron oxide contained in the chromite ore, or that which is allowed to penetrate the brick, can cause excessive expansion of the lattice because of the unequal diffusion of the iron and chromium ions in magnesia chrome spinel, which leads to the production of pores in the iron-brick phase.

Reactions Between Refractories and Gases. Reactions of refractories and gases can be quite destructive. The gases generally penetrate the pores of the refractory destroying its structure. The refractory may either expand and crack because of the formation of new, low density compounds, or its refractoriness may be drastically reduced because low melting compounds are formed. An example is the disintegration of aluminosilicates in blast furnaces caused by carbon monoxide. The deposition of carbon is catalyzed by iron in the bricks. The growth of the carbon deposit causes the brick to rupture and more surface is exposed. Therefore, a brick of low iron and alkali content having a dense, low permeability is preferred.

Reactions Between Refractories. In Table 17, the compatibilities of various refractories are given over a range of temperatures. Dissimilar refractories can react vigorously with each other at high temperatures. Phase diagrams are an excellent source of information concerning the reactivity between refractories.

Table 17. Approximate Initial Temperature, °C, at which Refractories React^{a, b}

Refractory	Magnesite, 92° MgO	Magnesite–chrome, CB ^c	Chrome–magnesite		Forsterite, stabilized	Chrome, fired	Alumina	
			CB ^c	Fired			90°	70°
magnesite, 92° MgO		>1700	>1700	>1700	(1700)	1700	>1700	1600
magnesite–chrome, CB ^c	>1700		>1700	>1700	1650	>1700	1450	1450
chrome–magnesite CB ^c	>1700	>1700		>1700	1650	>1700	1650	1600
fired	>1700	>1700	>1700		1650	>1700	1650	1650
forsterite, stabilized	(1700)	1650	1650	1650		1650	1650	1600
chrome refractories, fired	1700	>1700	>1700	>1700	1650		1650	1600
alumina 90°	>1700	1600	1600	1600	1650	1650		>1700
70°	1650	1600	1600	1500	1650	1600	>1700	
zircon	>1700	1600	1650	1600	1600	>1700	>1700	>1700
fireclay superduty	1400	(1700)	1650	1650	1600	1500	(1700)	(1700)
high duty	1400	1650	1600	1600	1650	1500	(1700)	(1700)
semisilica	1500	1400	1400	1500	1500	1500	(1500)	(1500)
silicon carbide, clay-bonded silica	1500	1400	1400	1400	1600	1500	1650	1650
type B ^d superduty	1500	1600	1600	1500	1650	1650	1650	1500
	1500	1600	1600	1600	1650	1650	1650	1500

^a Ref. 26; values in parentheses are estimated.
^b Max temperature tested 1700° C.
^c CB = chemically bonded (not fired); one or both materials not sufficiently refractory for test.
^d Ref. 34.

Silica Refractories. This type consists mainly of silica in three crystalline forms: cristobalite [14464-46-1], tridymite [1546-32-3], and quartz [14808-60-7]. Quartzite sands and silica gravels are the main raw materials, although lime and iron oxides are added to increase the mineralization of the tridymite and cristobalite. Uses include roof linings, refractories for coke ovens, coreless induction foundry furnaces, and fused-silica technical ceramic products. Consumption of silica refractories has declined dramatically since the 1960s as a result of the changes in the steel industry.

Fireclay Refractories. These products are made from clay minerals containing ca 17–45° Al₂O₃. Pure kaolin has the highest alumina content. Fireclay refractories are used in kilns, ladles, and heat regenerators, acid–slag-resistant applications, boilers, blast furnaces, and rotary kilns. They are generally inexpensive.

High Alumina Refractories. The desired alumina content, from 100° to just above 45°, is obtained by adding bauxites, synthetic aluminosilicates, and synthetic aluminas to clay and other bonding agents. These refractories are used in kilns, ladles, and furnaces that operate at temperatures or under conditions for which fireclay refractories are not suited. Phosphate-bonded alumina bricks have exceptionally high strength at low to intermediate temperatures and are employed in aluminum furnaces. High alumina and mullite are used in furnace roofs and petrochemical applications.

Chrome Refractories. Naturally occurring chrome ore composed mainly of chromite [53293-42-8] may be made into a brick or blended with fine calcined magnesite to obtain the desired chrome-to-magnesia ratio. When blended with magnesite, the products containing more chrome than magnesia are denoted chrome magnesite. These refractories are used in nonferrous metallurgical furnaces, rotary kiln linings, and secondary refining vessels, such as argon–oxygen decarborizers (AODs) and glass-tank regenerators. Chrome and alumina are isomorphous and form a complete series of solid solutions. Small amounts of chrome in alumina increase the resistance to metal penetration dramatically.

Magnesite Refractories. These refractories do not contain magnesite as their name implies, but rather periclase. The term magnesite refers to the ore from which the periclase was made, although magnesite brick can be made from synthetic periclase, ie, seawater MgO. Shaped magnesite refractories may be impregnated or bonded with pitch or resin to improve resistance to slag attack. Chrome ore can be added to magnesite to produce magnesite–chrome refractories used in lining and maintenance of steelmaking and refining vessels and checkers.

Dolomite Refractories. Refractories containing dead-burned dolomite and possibly fluxes such as millscale, serpentine, or clay are dolomite refractories. Shaped refractories may be bonded or impregnated with pitch to improve slag resistance and inhibit hydration. Addition of magnesite gives magnesite–dolomite (magdol) refractories. Dolomite refractories are primarily used in linings of BOF vessels and refining vessels, and in ladles and cement kilns.

Spinel Refractories. These refractories, which contain synthetic spinel, MgAl₂O₄, exhibit good strength at high temperatures as well as thermal shock resistance. Chromium and magnesium oxides crystallize in the same structure and are referred to as chrome–magnesite spinels. Spinel refractories have been used in cement kilns, AOD vessels, and steel-ladle linings.

Forsterite Refractories. Refractories made from forsterite, Mg₂SiO₄, resist alkali attack and have good volume stability, high temperature strength, and fair resistance to basic slags. Uses include nonferrous metal furnace roofs and glass-tank refractories not in contact with the melt, ie, checkers, ports, and uptakes.

Silicon Carbide Refractories. Silicon carbide has a wide range of refractory uses including chemical tanks and drains, kiln furniture, abrasion-resistant linings, blast-furnace linings, and nonferrous metallurgical crucibles and furnace linings. These materials are also used in power plant cyclone boilers and in municipal waste-to-energy incineration (see INCINERATORS; POWER GENERATION).

Zirconia Refractories. The most common zirconia-containing refractories are made from zircon sand and are used mostly for container glass-tank subpaver brick. Refractory blocks made from a composition of zircon and alumina, used to contain glass melts, are generally electromelted and then cast. These exhibit excellent corrosion resistance but are subject to thermal shock. Refractories made from pure ZrO₂ are extremely expensive and are reserved for extra high temperature service above 1900°C. Additives such as yttria or CaO and MgO prevent deterioration during heating and cooling.

BIBLIOGRAPHY

"Refractories" in *ECT* 1st ed., Vol. 11, pp. 597–433, by L. J. Trostel and R. P. Heuer, General Refractories Co.; in *ECT* 2nd ed., Vol. 17, pp. 227–267, by W. T. Bakker, G. D. Mackenzie, G. A. Russell, Jr., and W. S. Treffner, General Refractories Co.

1. A. F. Greaves-Walker, *Bull. Am. Ceram. Soc.* **6**, 20, 213 (1941).
2. F. Singer and S. S. Singer, *Industrial Ceramics*, Chemical Publishing Co., Inc., New York, 1964.
3. R. D. Pehlke and co-eds., *Basic Oxygen Furnace Steelmaking*, Vol. 4, The Iron and Steel Society of AIME, New York, 1977, Chapt. 11, pp. 1–58.
4. F. H. Norton, *Refractories*, 3rd ed., McGraw-Hill Co., Inc., New York, 1949.
5. J. J. Suec and co-eds., *Ceramic Data Book 1981 Suppliers' Catalog and Buyers' Directory*, Cahners Publishing Co., Denver, Colo., 1981.
6. W. D. Kingery, *Introduction to Ceramics*, John Wiley & Sons, Inc., New York, 1960.
7. A. Alper, ed., *High Temperature Oxides (1–4)*, Vol. 5, *Refractory Materials*, Academic Press, Inc., New York, 1970.
8. E. Ryshkewitch, *Oxide Ceramics*, Academic Press, Inc., New York, 1960.
9. H. Salmang, *Ceramics, Physical and Chemical Fundamentals*, Butterworth and Co., Ltd., London, 1961.
10. J. R. Hague, J. F. Lynch, A. Rudnick, F. C. Holden, and W. H. Duckworth, eds., *Refractory Ceramics for Aerospace: A Materials Selection Handbook*, The American Ceramic Society, Inc., Columbus, Ohio, 1964.
11. B. Phillips, *Res. Dev.* **18**, 22 (1967).
12. E. M. Levin, C. R. Robbins, and H. F. McMurdie, *Phase Diagrams for Ceramists*, The American Ceramic Society, Inc., Columbus, Ohio, 1964.
13. L. A. Dahl, "Rock Products, Sept. to Dec., 1938," *PCA Res. Bull.* **1** (1939).
14. L. A. Dahl, *J. Phys. Chem.* **52**, 698 (1948).
15. C. N. Fenner, *Am. J. Sci.* **36**(4), 383 (1913).
16. J. W. Greig, *Am. J. Sci.* **13**(5), 1 (1927).
17. A. Muan and E. F. Osborn, *Phase Equilibria Among Oxides in Steelmaking*, Addison-Wesley Publishing Co., Inc., Reading, Mass., 1965.
18. N. L. Bowen and J. F. Schairer, *Am. J. Sci.* **29**(5), 153 (1935).
19. P. Duwez, F. Odell, and F. H. Brown, Jr., *J. Am. Ceram. Soc.* **35**(5), 109 (1952).
20. H. E. McGonnon, ed., *The Marking, Shaping and Treating of Steel*, 9th ed., U.S. Steel Corp., Pittsburgh, Pa., 1970.
21. *Refractories for Industry*, C-E Refractories, Valley Forge, Pa., 1976.
22. Technical data, Kaiser Refractories, Oakland, Calif., 1977.
23. *Criterion I and Other Class Furnace Refractories*, C-E Refractories, Valley Forge, Pa., 1977.
24. C. O. Fairchild and M. F. Peters, *J. Am. Ceram. Soc.* **9**, 700 (1926).
25. *Fibrous Ceramic Thermal Insulation for Ultra-High Temperature Use*, Zircar Products, Inc., New York, 1976.
26. *Modern Refractories Practice*, Harbison-Walker Refractories Co., Pittsburgh, Pa., 1961.
27. P. P. Budnikov, *The Technology of Ceramics and Refractories*, The MIT Press, Cambridge, Mass., 1964.
28. J. E. Neal and R. S. Clark, *Chem. Eng.*, 56 (May 4, 1981).
29. *Refractories*, General Refractories Co., Philadelphia, Pa., 1949.
30. A. A. Litvakovski, *Fused Cast Refractories*, Israel Program for Scientific Translations, Jerusalem, 1961.
31. *Refractories*, Current Industrial Reports, U.S. Bureau of Census, Industry Division, Annual Reports, Washington, D.C., 1965–1979.
32. *Manual of ASTM Standards on Refractory Materials*, 8th ed., American Society for Testing and Materials, Philadelphia, Pa., 1957.
33. E. Criado, A. Pastor, and R. Sancho, *Am. Ceram. Soc.*, 169 (1989).
34. "Refractories, Glass, Ceramic Materials; Carbon and Graphite Products," *ASTM Annual Book of ASTM Standards*, Vol. 15.01, ASTM, Philadelphia, Pa., 1992.
35. J. F. Burst and J. A. Spieckerman, *Chem. Eng.*, 85 (July 31, 1967).
36. *Industrial Surveys of Refractory Service Conditions*, Committee C-8, ASTM, Philadelphia, Pa.

General References

L. J. Trostel, ed., *Proceedings of the 1989 Unified International Technical Conference on Refractories (Unitect)*, American Ceramic Society, Columbus, Ohio, 1989.

S. C. Carniglia and G. L. Barna, *Handbook of Industrial Refractories Technology—Principles, Types, Properties and Applications*, Noyes Publications, Park Ridge, N.J., 1992.

G. L. Barrows, S. H. Chen, and L. Shemanski, *Bull. Am. Ceram. Soc.* **72**(7), 28–34 (1993).

P. A. Janeway, ed., *Bull. Am. Ceram. Soc.* **73**(10), 46–55 (1994).

J. R. Rait, *Basic Refractories, Their Chemistry and Their Performance*, Ilife & Sons, Ltd., London, 1950.

J. H. Chesters, *Steel Plant Refractories, Testing Research and Development*, The United Steel Co., Ltd., Sheffield, U.K., 1963.

J. R. Coxey, *Refractories*, The Pennsylvania State College, State College, Pa., 1950.

L. R. McCreight, H. W. Rauch, and W. H. Sulton, eds., *Ceramic and Graphite Fibers and Whiskers*, Vol. 1, *Refractory Materials*, Academic Press, Inc., New York, 1970.

E. K. Storms, ed., *The Refractory Carbides*, Vol. 2, *Refractory Materials*, Academic Press, Inc., New York, 1970.

A. M. Alper, ed., *Phase Diagrams, Materials Science and Technology (1–3)*, Vol. 6, *Refractory Materials*, Academic Press, Inc., New York, 1970.

A. K. Kulkarni and V. K. Moorthy, eds., *Proceedings of the 1st Symposium on Material Science Research*, Series 2, Chemical Metallurgy Commission, Dept. of Atomic Energy, Bombay, India, 1970, pp. 208–218.

G. R. Belton, ed., *Proceedings of the International Conference of Metal Material Science, 1969*, Plenum Press, New York, 1970.

H. Bibring, G. Seibel, and M. Rabinouitch, eds., *Proceedings of the 2nd International Conference of Strength Metals Alloys*, Series 3, ASM, Metals Park, Ohio, 1970, pp. 1178–1182.

G. H. Criss and A. R. Olsen, eds., *Proceedings of the 3rd National Incinerator Conference*, American Society of Mechanical Engineering, New York, 1968, pp. 53–68.

C. Brosset, ed., *Trans. Int. Ceram. Congr.*, 10 (1967).

R. M. Fulrath and J. A. Pask, eds., *Ceramic Microstructures, Proceedings of the 3rd International Material Symposium*, John Wiley & Sons, Inc., New York, 1968.

G. C. Kuczynski, ed., *Sintering Related Phenomena, Proceedings of the 2nd International Conference*, Gordon and Breach Science Publishers, New York, 1967.

H. H. Hausner, ed., *Fundamental Refractory Compounds*, Plenum Press, New York, 1968.

R. C. Bradt, D. P. H. Hasselman, and F. F. Lange, eds., *Mechanical Ceramics, Proceedings of the Symposium*, Plenum Press, New York, 1974.

S. J. Lefond, *Industrial Mineral Rocks*, 4th ed., American Institute of Mechanical Engineers, New York, 1975.

C. S. Tedmon, Jr., *Corrosion Problems in Energy Conversion Generators*, Electrochemical Society, Princeton, N.J., 1974.

J. J. Burke, A. E. Gorum, and N. R. Katz, eds., *Ceramic High-Performance Applications, Proceedings of the 2nd Army Materials Technology Conference*, Brook Hill Publishing Co., Chestnut Hill, Mass., 1974.

R. C. Bradt and R. E. Tressler, eds., *Deformation of Ceramic Materials, Proceedings of the 1974 Symposium*, Plenum Press, New York, 1975.

R. F. S. Fleming, ed., *Proceedings of the Industrial Mineral International Congress*, Metallurgical Bulletin Ltd., London, 1975.

Z. A. Foroulis and W. W. Smeltzer, eds., *Metallurgical Slag—Gas Reaction Processes*, Electrochemical Society, Inc., Princeton, N.J., 1975.

S. Modry and M. Svata, eds., *Pore Structures, Properties, and Materials, Proceedings of the International Symposium*, Scademia, Prague, Czechoslovakia, 1974.

F. V. Tooley, *Handbook of Glass Manufacturing*, Books Industries, Inc., New York, 1974.

10th International Congress of Glass, Ceramic Society of Japan, Tokyo, 1974.

N. Standish, ed., *Alkalies Blast Furnaces, Proceedings of the 1973 Symposium*, Dept. of Metallurgical Material Science, McMaster University, Hamilton, Ontario,

Canada, 1973.

R. C. Marshall, ed., *Silicon Carbide, Proceedings of the 3rd International Conference, 1973*, University of South Carolina Press, Columbia, S.C., 1974.

B. Cockayne, ed., *Modern Oxide Materials, Prep., Prop. Device Applications*, Academic Press, London, 1972.

R. R. M. Johnston, ed., *Corrosion Technology in the Seventies, 12th Annual Conference of the Australasian Corrosion Association*, Parkville, Victoria, Australia, 1972.

L. D. Pye, ed., *Introduction to Glass Science, Proceedings of a Tutorial Symposium*, Plenum Press, New York, 1972. Review of glass melting refractory corrosion.

High Temperature Material, Proceedings of the 3rd Symposium on Material Science Research, Dept. of Atomic Energy, Bombay, India, 1972. Papers cover a variety of oxide and nonoxide high temperature materials like spinel, alumina, and silicon nitride.

J. J. Burke, ed., *Powder Metal High-Performance Applications, Proceedings of the 18th Sagamore Army Material Research Conference*, Syracuse University Press, Syracuse, N.Y., 1972. Review on silicon carbide–silicon nitride ceramics.

Mechanical Behavior of Materials, Proceedings of the 1st International Conference, Society of Material Science, Kyoto, Japan, 1972.

J. D. Buckley, ed., *Advanced Materials, Composite Carbon, Preparation Symposium*, American Ceramics Society, Inc., Columbus, Ohio, 1972. Papers on composite materials including carbon and graphite or nitride composites.

J. I. Duffy, "Refractory Materials, Developments Since 1977," *Chemical Technology Review*, No. 178, Noyes Data Corp., Park Ridge, N.J., 1980.

I. Ahmad and B. R. Noton, eds., *Advanced Fibers Compositions at Elevated Temperatures, Proceedings of the Symposium of the Metallurgical Society*, American Institute of Mechanical Engineers, Warrendale, Pa., 1980.

G. V. Samsonov and I. M. Vinitskii, *Handbook of Refractory Compounds*, Plenum Press, New York, 1980.

A. V. Levy, ed., *Proceedings of the Corrosion/ Erosion Coal Conversion Systems Materials Conference*, National Association of Corrosion Engineers, Houston, Tex., 1979.

Basic Oxygen Steelmaking: New Technology Emerges? Proceedings of the Conference, Metallurgical Society of London, 1979. Contains papers on refractories for conventional and new bottom-blown vessels.

M. R. Louthan, Jr. and R. P. McNitt, eds., *Environmental Degradation of Engineering Materials, Proceedings of the Conference*, Virginia Polytechnical Institute, Blacksburg, Va., 1978.

19th International Refractories Colloquium, Institute Gesteinshuttenkunde RWTH Aachen, Aachen, Germany, 1976. Contains a number of good papers covering a wide range of refractory materials, properties, and problems.

Contain. Cast. Steel, Proceedings of the International Conference, Metallurgical Society of London, 1977.

3rd U.S./U.S.S.R. Colloquium on Magnetohydrodynamic Electric Power Generation, National Technical Information Service, Springfield, Va., 1976.

P. Vincenzini, ed., *Advanced Ceramic Processes, Proceedings of the 3rd International Meeting of Modern Ceramics Technology*, National Research Council, Research Laboratory of Ceramics Technology, Faenza, Italy, 1978.

G. Y. Onoda and L. L. Hench, eds., *Ceramic Processing Before Firing*, John Wiley & Sons, Inc., New York, 1978.

V. I. Matkovich, ed., *Boron Refractory Borides*, Springer-Verlag, Berlin, 1977.

R. M. Fulrath and J. A. Pask, *Ceramic Microstructures: Proceedings of the 6th International Materials Symposium*, Westview Press, Boulder, Colo., 1977.

H. David Leigh III
Clemson University

REFRACTORY COATINGS

Refractory coatings denote metallic, refractory compound, ie, oxides, carbides (qv), and nitrides (qv), and metal-ceramic coatings associated with high temperature service as contrasted to coatings (qv) used for decorative or corrosion-resistant applications. Coatings of high melting materials that are used in other than high temperature applications are also considered to be refractory. A coating may be defined as a near-surface region having properties that differ significantly from the bulk of the substrate (see CERAMICS; METALLIC COATINGS; METAL SURFACE TREATMENTS).

The highest melting refractory metals are tungsten, tantalum, molybdenum, and niobium, although titanium, hafnium, zirconium, chromium, vanadium, platinum, rhodium, ruthenium, iridium, osmium, and rhenium may be included (see REFRACTORIES). Many of these metals do not resist air oxidation. Hence, very few, if any, are used in elemental form for high temperature protection. However, bulk alloys based on nickel, iron, and cobalt and alloying elements such as chromium, titanium, aluminum, vanadium, tantalum, molybdenum, silicon, and tungsten are used extensively in high temperature service. Some modern high temperature oxidation- and corrosion-resistant coatings have compositions similar to the high temperature bulk alloys (see HIGH TEMPERATURE ALLOYS) and are applied by thermal spraying, evaporation, or sputtering. The protection mechanism for these high temperature alloy coatings is based on adherent impervious surface films of Al₂O₃, SiO₂, CrO₂, or a spinel-type material that grow upon high temperature exposure to air.

Refractory coatings also include materials having high melting points, eg, silicides, borides, carbides, nitrides, or oxides, and combinations such as oxycarbides, etc. In addition, mixtures of metals and refractory compounds, sometimes called metallides, of various microstructural configurations, ie, laminates, dispersed phases, etc, can also be classified as refractory coatings. Other terms used to describe such multiphase coatings are multilayer coating, multicoating, composite coating, reinforced coating, and the like.

In some cases, a coating is a new material that is deposited onto the substrate by a variety of methods. It is then called a deposited or overlay coating. In other cases, the coating may be produced by altering the surface material to produce a surface layer composed of both the added and substrate materials. This is called a conversion coating, cementation coating, diffusion coating, or chemical-conversion coating when chemical changes in the surface are involved. Coatings may also be formed by altering the properties of the surface by melting and quenching, mechanical deformation, or other processes that change the properties without changing the composition.

Coating technology has developed extensively since the 1980s. In this article, emphasis is given to those processes which have proven to be commercially viable or to processes that have undergone extensive development in the 1990s.

All coating methods consist of three basic steps: synthesis or generation of the coating species or precursor at the source, transport from the source to the substrate, and nucleation, growth, or buildup of the coating on the substrate. These steps can be completely independent of each other or may be superimposed on each other, depending on the coating process (see COATING PROCESSES). A process in which the steps can be both varied independently and controlled offers great flexibility, enabling a larger variety of materials to be deposited.

Numerous schemes can be devised to classify deposition processes. The scheme used herein is based on the dimensions of the depositing species, ie, atoms and molecules, softened particles, liquid droplets, bulk quantities, or the use of a surface-modification process (1,2). Coating methods are as follow (2):

Atomistic deposition	Bulk coatings
electrolytic environment	wetting processes
electroplating	painting
electroless plating	dip coating
fused-salt electrolysis	electrostatic spraying
chemical displacement	printing
vacuum environment	spin coating
vacuum evaporation	cladding
ion beam deposition	explosive
molecular beam epitaxy	roll bonding
plasma environment	overlaying
sputter deposition	weld coating
activated reactive evaporation	liquid-phase epitaxy
plasma polymerization	
ion plating	
chemical vapor environment	
chemical vapor deposition	
reduction	
decomposition	
plasma enhanced	
spray pyrolysis	
Particulate deposition	Surface modification
thermal spraying	chemical conversion
plasma spraying (at 50 kW)	
high power plasma spraying (at 250 kW)	electrolytic
D-gun	anodization (oxides)
flame spraying	fused salts
arc wire	chemical–liquid
HVOF	chemical–vapor
fusion coatings	thermal
thick-film ink	plasma
enameling	leaching
electrophoretic	mechanical
impact plating	shot peening
	thermal
	surface enrichment diffusion from bulk
	sputtering
	ion implantation

The coating has to adhere to the substrate. In older thermal spray technology, the bonding may be mechanical as a result of the interlocking between

the asperities on the surface and the coating. In this case the surface roughness, ie, the average distance between asperities, must be equal to or larger than the dimensions of the depositing particles. These sprayed coatings internally do not adhere to a polished metal surface. In such cases the substrate surface has to be coarsened by techniques such as grit blasting that not only provide an anchor, but also clean the surface by removing oxide layers and scales, and provide high energy sites for a denser nucleation of coating crystallites and, to some extent, increase the real interfacial area. High velocity processes such as D-gun, high velocity oxy fuel (HVOF), and high powered plasma, eg, PlazJet (TAFA Inc. (Concord, New Hampshire)), propel particles at 4–10 times the impact velocity of other spray devices and provide coating bonds two to three times that of older spray processes, in many cases on clean, ground substrates (see PLASMA TECHNOLOGY). These particular processes also provide denser and in many cases pore-free coatings having thicknesses in the range of 325–1000 μm.

In diffusion or chemical bonding, the substrate and the coating material interdiffuse at the interface. The latter depends on the equilibrium between the two materials at the effective temperature of deposition. If the two materials have no solid solubility, a sharp or abrupt interface forms. The bond strength depends on the affinity between the materials and increases with the degree of wetting of the substrate with the coated material. If the two materials exhibit extensive solid solubility, the interface is a solid solution. If intermetallic phases are indicated on the equilibrium diagram, these may also form at the interface as may gas–metal compounds, depending on the degree of contamination present on the substrate or arriving at the substrate from the environment.

Residual stresses, which are always present in coatings, arise from thermal expansion mismatch between coating and substrate, or growth stresses caused by imperfections in the coating that are built-in during the process of film growth. Growth stresses are usually significant for coating processes carried out near room temperature. Failure or debonding of the coating may be caused by the residual stresses in the coating and substrate. The location of the failure may be in the substrate, at the interface, or in the coating, depending on plastic deformation and resistance to nucleation and growth of cracks, ie, fracture toughness. For example, a coating may fail at the interface because of crack propagation caused by a brittle intermetallic phase. A post-failure surface analysis of the substrate and the coating is recommended; the use of surface chemical analysis techniques, eg, scanning electron microscopy (sem) having energy-dispersive x-ray analysis (edax) capabilities, Auger spectroscopy, or esca analysis can determine the chemical composition at the failed interfaces and deduce its cause (see SURFACE AND INTERFACE ANALYSIS (SUPPLEMENT)).

Newer high velocity thermal spray coating processes produce coatings in compression rather than tension because of the shot peening effect of the supersonic particles on impact. This has permitted coating as thick as 12,500 μm without delamination as compared to older processes limited to 1,250 μm. The reduced residence time of particles at temperature minimizes decomposition of carbides present in conventional d-c plasma. This improves wear and hardness (qv) properties.

Coatings may be permeable either to the atmosphere or the substrate material. Diffusion of oxygen through a coating can result in gaseous products that may rupture the coating. Even if the gas can escape through the substrate, as for graphite, the substrate can be consumed and the bond weakened. Coatings permeable to the substrate material by outward diffusion can suffer similarly, as oxidation of the diffused species may occur at the atmosphere–coating interface.

The design of a coating that might equilibrate with its substrate during use is based on the phase diagram of the system. Extensive regions of solid solubility may indicate that rapid interdiffusion can be expected. For multicomponent coatings on multicomponent substrates, a large number of possibilities exists for formation of intermetallic compounds. Limited compound formation at the coating–substrate interface may be preferred if it decreases the permeability of the coating. Wetting or, more strictly speaking, high affinity between coating and substrate is conducive to high temperature service.

A reasonably close match of thermal expansion of the coating and substrate over a wide temperature range to limit failure caused by residual stresses is desired for coatings. Because temperature gradients cause stress even in a well-matched system, the mechanical properties, strength, and ductility of the coating as well as the interfacial strength must be considered.

The simple case of protecting tungsten or molybdenum by using platinum illustrates the above point. A wide range of solid solubility exists that contributes to interdiffusion between coating and base metal to form brittle intermetallic compounds. Both solutions and compounds oxidize preferentially, losing tungsten or molybdenum. Platinum is also somewhat permeable to oxygen, which permits subcoating oxidation. A thermal expansion mismatch exists between platinum and these intermetallic compounds and adds another complication where thermal cycling is required. Silicides and beryllides exhibit sufficient conductivity to be electroplated and could well be used as diffusion barriers. However, the bond strengths and mechanical compatibility of two layers instead of one must be considered.

Atomic Deposition Processes

Electrodeposition.

Aqueous Electrodeposition. The theory of electrodeposition is well known (see ELECTROPLATING). Of the numerous metals used in electrodeposition, only 10 have been reduced to large-scale commercial practice. The most commonly plated metals are chromium, nickel, copper, zinc, rhodium, silver, cadmium, tin, and gold, followed by the less frequently plated metals iron, cesium, platinum, and palladium, and the infrequently plated metals iridium, ruthenium, and rhenium. Of these, only platinum, rhodium, iridium, and rhenium are refractory.

The electrodeposition of tungsten alloys of iron, nickel, and cobalt is commercially feasible but has remained largely experimental. The properties of these alloys should, however, be of sufficient interest for engineering applications.

Cermets, ie, materials containing both ceramic and metal, eg, TiC–Ni and Al₂O₃–Cr, can be deposited from plating baths if the particulate matter is suspended by air agitation or stirring (see CERAMICS; COMPOSITE MATERIALS, CERAMIC MATRIX). Particle sizes from 1 to 50 μm may be deposited to concentrations over 20% . Chromium-based cermets containing zirconium and tungsten borides, zirconium nitride, and molybdenum carbide can be plated on refractory metals and graphite. However, at 1650°C, protection is afforded only for a few minutes. Cermets also suffer from porosity. Therefore, such coatings find application where exposure times are short and erosion conditions severe, and where they can be used to bridge the expansion mismatch between a metallic substrate and a ceramic coating.

Fused-Salt Electrodeposition. Molten salt electrolysis has been used since Davy's and Faraday's pioneering experiments, but application as a method for coating metals has remained of academic interest. Fused-salt baths may be used to plate ruthenium, platinum, and iridium with improved coating soundness. Developments in this technology may lead to commercial plating of other refractory metals, such as zirconium, hafnium, vanadium, niobium, tantalum, molybdenum, and tungsten. Molten salt electrolysis is based on the same principles as aqueous electrolysis processes, but the vehicle, ie, the bath, is a molten salt. The coating material deposits as a layer on the substrate.

Electroless Deposition. Electroless plating (qv) is defined as a controlled, autocatalytic chemical reduction process for depositing metals. It resembles electroplating because it can be run continuously to build up a thick coating. Unlike displacement processes, it does not involve a chemical reaction with the substrate metal; and unlike the well-known silver reduction process (Tollens reaction) for the silvering of optical glass, it is selective, ie, deposits form only on a catalytic surface. The metals nickel, cobalt, platinum, palladium, and gold, and nickel–cobalt alloys can be deposited. Chromium, iron, and vanadium are often claimed, but electroless plating of active metals must be viewed with suspicion. Most such processes are of the displacement type.

Physical Vapor Deposition Processes. The three physical vapor deposition (PVD) processes are evaporation, ion plating, and sputtering (see THIN FILMS, FILM FORMATION TECHNIQUES).

The materials deposited by PVD techniques include metals, semiconductors (qv), alloys, intermetallic compounds, refractory compounds, ie, oxides, carbides, nitrides, borides, etc, and mixtures thereof. The source material must be pure and free of gases and inclusions, otherwise spitting may occur.

Metals and Elemental Semiconductors. Evaporation of single elements can be carried out from various evaporation sources subject to the restrictions with regard to melting point, container reactions, deposition rate, etc. A typical arrangement is shown in Figure 1 for electron beam heating (3). This type of source is ideal for refractory coatings, because the material to be deposited is contained in a noncontaminating water-cooled copper crucible and the surface of the material can easily be heated to >3000° C.

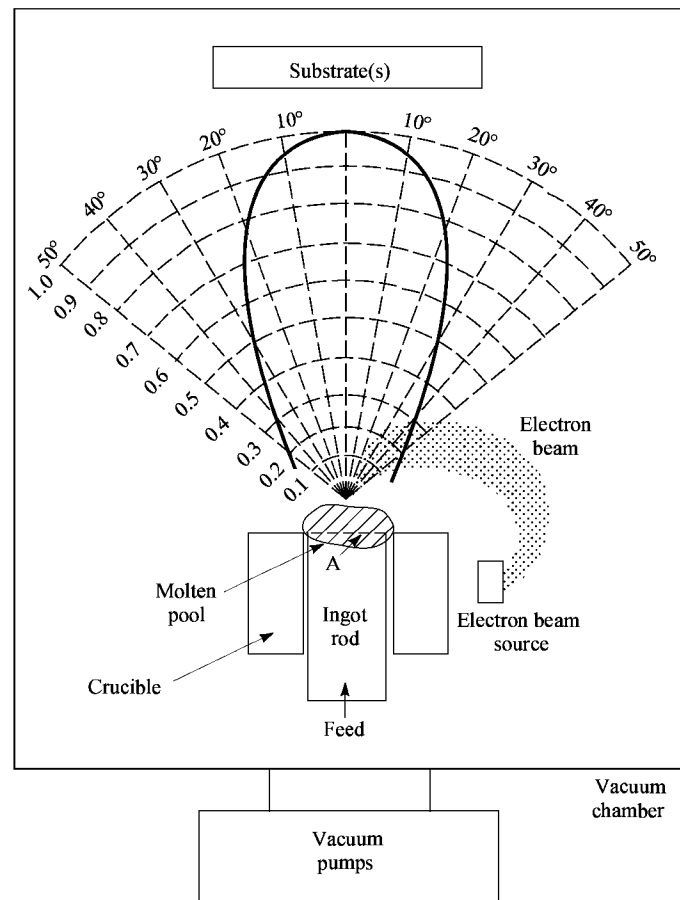


Fig. 1. Vacuum evaporation process with use of electron beam heating where A represents the material to be deposited. The flux profile (—) is at a range of 1.0 m ± 50° from perpendicular.

Alloys. Alloys consist of two or more elements of different vapor pressures and hence different evaporation rates. As a result, the vapor phase and therefore the deposit constantly vary in compositions. This problem can be solved by multiple sources or a single rod- or wire-fed electron beam source fed with the alloy. These solutions apply equally to evaporation or ion-plating processes.

Multiple Sources. Multiple sources offer a more versatile system. The number of sources evaporating simultaneously is equal to or less than the number of constituents in the alloy. The material evaporated from each source can be a metal, alloy, or compound. Thus, it is possible to synthesize a dispersion strengthened alloy, eg, Ni–ThO₂. However, the evaporation rate from each source has to be monitored and controlled separately. The source-to-substrate distance would have to be sufficiently large (38 cm for 5 cm dia sources) to blend the vapor streams prior to deposition, which decreases the deposition rate (Fig. 2). Moreover, if the density of two vapors differs greatly, it may be difficult to obtain a uniform composition across the width of the substrate because of scattering of the lighter vapor atoms. Evaporating each component sequentially produces a multilayered deposit that is homogenized by annealing. High deposition rates are difficult to obtain.

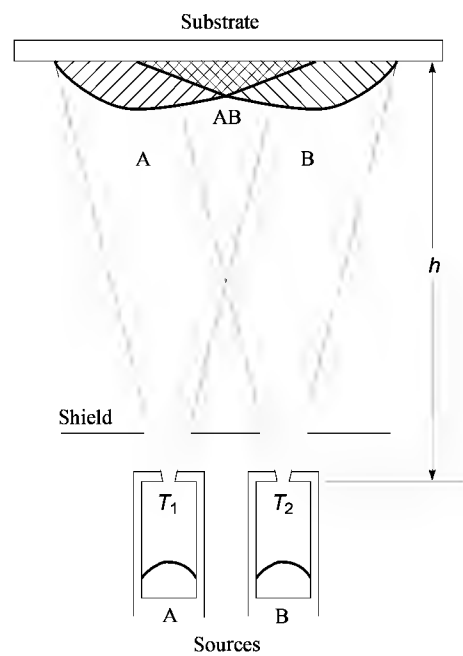


Fig. 2. Two-source evaporation at temperatures T_1 and T_2 defining the regions coated by materials A, B, and AB, where h represents height.

Single Rod-Fed Electron Beam Source. The disadvantages of multiple sources for alloy deposition can be avoided by using a single wire-fed or rod-fed source (Fig. 3) (3). A molten pool of limited depth is above the solid rod. If the equilibrium vapor pressures of the components of an alloy A_1B_1 are in the ratio of 10:1 and the composition of the molten pool is $A_{10}B_1$, under steady-state conditions, the composition of the vapor is the same as that of the solid being fed into the molten pool. The procedure can be started with a pellet of appropriate composition $A_{10}B_1$ on top of a rod A_1B_1 to form the molten pool initially, or with a rod of alloy A_1B_1 to evaporate the molten pool until it reaches composition $A_{10}B_1$. The temperature and volume of

the molten pool must be constant to obtain a constant vapor composition. A theoretical model has been developed and confirmed by experiment, and deposits of Ni–20 wt % Cr, Ti–6 wt % Al–4 wt % V, Ag–5 wt % Cu, Ag–10 wt % Cu, Ag–20 wt % Cu, Ag–30 wt % Cu, and Ni–Cr–Al–Y alloy have been successfully prepared. This method can be used with a 5000-fold vapor pressure difference between components. It cannot be used when one of the alloy constituents is a compound, eg, Ni–ThO₂.

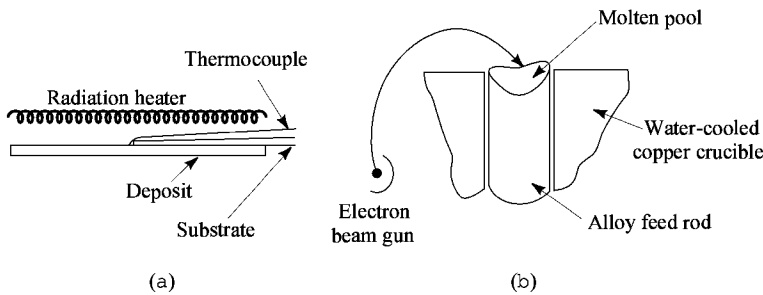


Fig. 3. Alloy evaporation from a single rod-fed source under steady-state conditions: $p^\circ = 10p^\circ_{AB}$; feed rod, A_1B_1 ; molten pool, $A_{10}B_1$; and vapor and deposit, A_1B_1 , where p° the equilibrium vapor pressure of component B B, and p° the equilibrium vapor pressure of component A A. Part (a) shows wire-fed, and (b) rod-fed sources.

Sputtering. Sputtering deposits alloys by means of an alloy target. The surface composition of the target changes in the inverse ratio of the sputtering yields of the individual elements, as in the alloy evaporation from a single rod-fed electron beam source. Alternatively, the sputtering target can be made of strips of the components of the alloy with the respective surface areas inversely proportional to their sputtering yield.

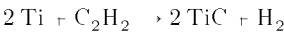
Refractory Compounds. Refractory compounds resemble oxides, carbides, nitrides, borides, and sulfides in that they have a very high melting point. In some cases, they form extensive defect structures, ie, they exist over a wide stoichiometric range. For example, in TiC, the C:Ti ratio can vary from 0.5 to 1.0, which demonstrates a wide range of vacant carbon lattice sites.

In direction evaporation, the evaporant is the refractory compound itself, whereas in reactive or activated reactive evaporation (ARE), a metal or a low valency metal compound is evaporated in the presence of a partial pressure of a reactive gas to form a compound deposit, eg, Ti is evaporated in the presence of N₂ to form TiN, or Si or SiO is evaporated in the presence of O₂ to form SiO₂.

Direct Evaporation. Evaporation can occur with or without dissociation of the compound into fragments. The observed vapor species show that very few compounds evaporate without dissociation. Examples are MgF₂, B₂O₃, CaF₂, SiO, and other Group 14 (IV) divalent oxides, eg, SiO homologues such as GeO and SnO.

In general, when a compound is evaporated or sputtered, the material is not transformed to the vapor state as compound molecules but as fragments thereof. Subsequently, the fragments recombine, most probably on the substrate, to reconstitute the compound. Therefore, the stoichiometry (anion:cation ratio) of the deposit depends on the deposition rate, the ratios of the various molecular fragments, the impingement of other gases present in the environment, the surface mobility of the fragments which in turn depends on their kinetic energy and substrate temperature, the mean residence time of the fragments of the substrate, the reaction rate of the fragments on the substrate to reconstitute the compound, and the impurities present on the substrate. For example, direct evaporation of Al₂O₃ results in a deposit deficient in oxygen, ie, having the composition Al₂O_{3-x}. This O₂ deficiency could be compensated for by introducing O₂ at a low partial pressure into the environment.

Reactive Evaporation. In reactive evaporation (RE), metal or alloy vapors are produced in the presence of a partial pressure of reactive gas to form a compound either in the gas phase or on the substrate as a result of a reaction between the metal vapor and the gas atoms:



If the metal and gas atoms are activated or ionized in the vapor phase, which activates the reaction, the process is called the activated reactive process (ARE), as illustrated in Figure 4 (4). The metal is heated and melted by a high acceleration voltage electron beam. The melt has a thin plasma sheath on top from which low energy secondary electrons are pulled upward into the reaction zone by an electrode placed above the pool. The electrode is biased to a low positive d-c potential (20–100 V). These low energy electrons have a high ionization cross section, thus ionizing or activating the metal and gas atoms and increasing the reaction probability on collision. Titanium carbide was synthesized with this process by reaction of Ti metal vapor and C₂H₂ gas having a carbon:metal ratio approaching unity. Moreover, by varying the partial pressure of either reactant, the carbon:metal ratio of carbides could be varied at will. This process has also been applied to the synthesis of the five different Ti oxides (5). Using the ARE process, ie, with a plasma, as compared to the RE process, ie, without a plasma, a higher oxide formed for the same partial pressure of O₂, which demonstrates a better gas utilization in the presence of plasma (see PLASMA TECHNOLOGY).

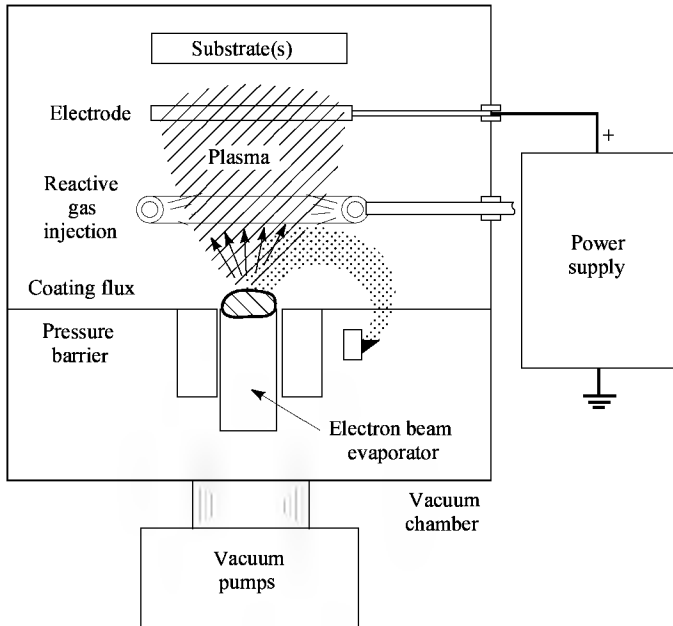


Fig. 4. Activated reactive evaporation process (4).

Reactive-Ion Plating. In reactive-ion plating (RIP), as in the reactive evaporation process, the metal atoms and reactive gases form a compound aided by the presence of a plasma. Because the partial pressures on the gases are much higher ($>1.3 \text{ Pa}$ (10^{-2} mm Hg)) than in the ARE process (13 mPa (10^{-4} mm Hg)), the deposits may be porous or sooty. In the simple diode ion-plating process, the plasma cannot be supported at a lower pressure. Therefore, an auxiliary electrode adjusted to a positive low voltage, as originally conceived for the ARE process, is used to initiate and sustain the plasma at a low pressure (ca 0.13 Pa (10^{-3} mm Hg)), as shown in Figure 5 (6).

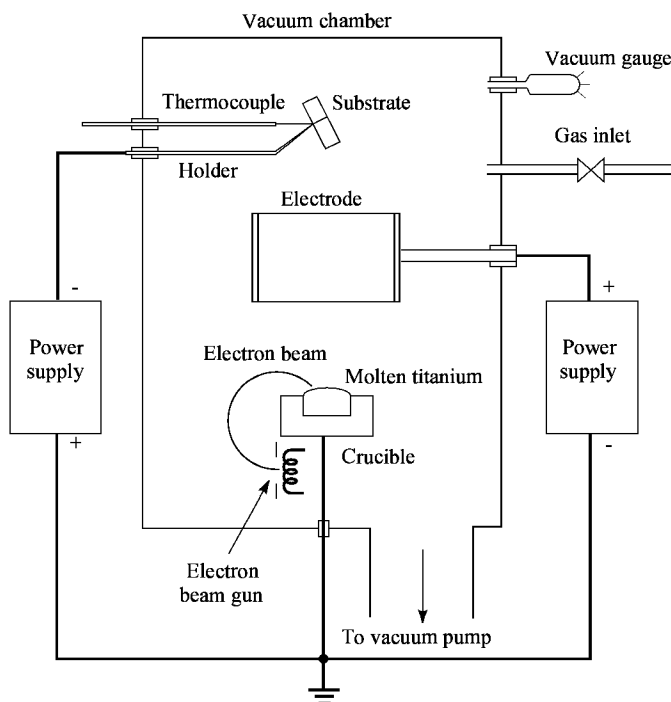


Fig. 5. Reactive-ion plating using auxiliary electrode for low pressure operation in deposition of compounds (6).

Courtesy of Kobayashi and Doi.

Reactive Sputtering. Reactive sputtering is similar to reactive evaporation and reactive-ion plating in that at least one coating species enters the system in the gas phase. Examples include sputtering Al in O_2 to form Al_2O_3 , Ti in O_2 to form TiO_2 , In-Sn in O_2 to form tin-doped In_2O_3 , Nb in N_2 to form NbN, Cd in H_2S to form CdS, In in PH_3 to form InP, and Pb-Nb-Zr-Fe-Bi-La in O_2 to form a ferroelectric oxide.

By reactive sputtering, many complex compounds can be formed from relatively easy-to-fabricate metal targets, insulating compounds can be deposited using a d-c power supply, and graded compositions can be formed, as described. The process, however, is complicated.

Chemical Vapor Deposition and Plasma-Assisted Chemical Vapor Deposition. In chemical vapor deposition (CVD), thin films or bulk coatings up to 2.5 cm in thickness are deposited by means of a chemical reaction between gaseous reactants passing over a substrate. Temperatures can be anywhere between 200 and 2200°C, but are usually between 500 and 1100°C. The optimum for a given reaction often lies within a very narrow range, and the process needs to be tailored to the substrate and the intended application. The substrate's melting point and susceptibility to chemical attack by the reacting gases or by their side products has to be considered. Coatings have application in a wide array of corrosion- and wear-resistant uses, but also in decorative layers, semiconductors, and magnetic and optical films (see THIN FILMS, FILM FORMATION TECHNIQUES).

In most cases, CVD reactions are activated thermally, but in some cases, notably in exothermic chemical transport reactions, the substrate temperature is held below that of the feed material to obtain deposition. Other means of activation are available (7), eg, deposition at lower substrate temperatures is obtained by electric-discharge plasma activation. In some cases, unique materials are produced by plasma-assisted CVD (PACVD), such as amorphous silicon from silane where 10–35 mol % hydrogen remains bonded in the solid deposit. Except for the problem of large amounts of energy consumption in its formation, this material is of interest for thin-film solar cells. Passivating films of SiO_2 or $\text{SiO}_2\text{-Si}_3\text{N}_4$ deposited by PACVD are of interest in the semiconductor industry (see SEMICONDUCTORS).

Plasma-assisted CVD processes use deposition temperatures lower than CVD processes. The desired deposition reaction is aided by the energy present in the plasma (8). The plasma greatly extends the utility of CVD processes, eg, the ability to deposit films on substrates that cannot withstand the temperature needed for the CVD process in reactions such as polymerization, anodization, nitriding, deposition of amorphous silicon, amorphous carbon (diamond-like carbon), etc. The films deposited by PACVD are generally more complex than their analogues deposited by normal CVD methods. For example, silicon nitride films have a composition $\text{Si}_3\text{N}_4\text{H}_x$ and may contain as much as 15 to 30 atomic % hydrogen as contrasted to the Si_3N_4 composition of the film deposited by the normal CVD process. Other advantages of PACVD are substrate surface etching and activation to produce good bonding at low deposition temperatures.

Ion Implantation. High energy ion implantation (qv) is a highly successful process for doping semiconductors to precisely controlled concentrations. The part's surface is bombarded with high energy ions ($\sim 50 \text{ keV}$), which results in a layer of implanted ions at an approximate depth of 8 nm. The potential for producing refractory coatings by implantation of specific metal ions or by reactions within the substrate to form refractory compounds exists but has not been exploited to date.

Particulate Deposition Processes

Thermal Spraying. In thermal spraying, ie, particles heated or melted, a combustion flame or electric arc heats particles of the refractory-coating material to a temperature sufficient to achieve sintering or cohesive solidification when the particles impinge on a substrate. Penetration of the substrate rarely accompanies the coalescence. The techniques include flame plating (D-gun, HVOF), flame spraying, arc wire, d-c plasma spraying, and induction-plasma spraying. These processes generate considerable noise, and acoustical protection is needed.

Flame plating (D-gun) employs oxygen and fuel gas. In this method, developed by the Union Carbide Corporation, the gas mixture is detonated by an electric spark at four detonations per second. The powders, mixed with the gas, are fed under control into a chamber from which they are ejected when detonation occurs. The molten, 14–16- μm particles are sprayed at a velocity of 732 m/s at distances of 5.1–10.2 cm from the surface. The substrate is moved past the stationary gun.

A steady-state rocket-type combustion spray unit has been developed, called high velocity oxy fuel (HVOF), that creates a steady state, continuous, supersonic spray stream (1.2–3 mm dia) resembling a rocket motor exhaust. The portable device injects and accelerates the particles inside a barrel (rocket nozzle). It produces coating quality and particle velocities equal to the D-gun at 5–10 times the spray rate with significantly reduced coating costs.

Flame spraying is no longer the most widely used melt-spraying process. In the power-feed method, powders of relatively uniform size (<44 μm (325 mesh)) are fed at a controlled rate into the flame. The torch, which can be held by hand, is aimed a few cm from the surface. The particles remain in the flame envelope until impingement. Particle velocity is typically 46 m/s, and the particles become at least partially molten. Upon impingement, the particles cool rapidly and solidify to form a relatively porous, but coherent, polycrystalline layer. In the rod-feed system, the flame impinges on the tip of a rod made of the material to be sprayed. As the rod becomes molten, droplets of material leave the rod with the flame. The rod is fed into the flame at a rate commensurate with melt removal. The torch is held at a distance of ca 8 cm from the object to be coated; particle velocities are ca 185 m/s.

Arc-plasma spraying equipment has been commercially available since 1958. This process does not utilize a fully ionized gas (plasma). Uniform particles (<74 μm (200 mesh)) are fed into the jet that emerges through the nozzle of a high pressure d-c arc. The particles are melted and ejected at controllable velocities of 30–210 m/s. The composition of the atmosphere can be varied over a wide range, ie, wet, dry, oxidizing, reducing (<80%), or inert. A 40–50 kW, 600 A d-c source supplies the average power requirements. The uv radiation generated during operations presents a safety hazard. During the early 1990s, high power (250 kW) d-c plasma spray systems (PlazJet) have been introduced. These can produce supersonic inert gas velocities, higher than those of HVOF, without combustion and the attendant oxygen. The coatings are significantly more dense, better bonded, and can be made to produce different, more desirable microstructures than conventional plasma.

Induction-plasma spraying is used in partial vacuum (ca 40 kPa (300 torr)) spraying of titanium aluminides for aerospace applications. A spark creates an ignition arc within the torch tube and coil to start the r-f power coupling to the levitated arc within the ceramic tube surrounded by the induction coil. This 5–8 cm egg-shaped plasma symmetrically fills the tube. Powders are fed into the plasma through a probe near the inlet of the plasma where the melted particle residence time can be varied over a greater range than in the d-c plasma process because gas velocity can be varied by adjusting the nozzle size and operating pressure. This apparatus is powered by a 100-kW, 450-kHz power source and cannot be held by hand. As in flame plating, the object is moved in the path of the gun.

Arc wire utilizes two continuously fed 1.6-mm dia intersecting wires with a d-c arc maintained between the wire tips as they meet. Compressed gas (usually air) strips the molten metal from the tips and forms a directional spray stream. This process is widely used to spray most metals. Arc wire is the most economical process because of the wire feedstock. Moreover, it utilizes ~10% of the thermal energy of the other spray processes (0.4 vs 6.6 kWh/kg using stainless steel) because of the direct arc heating of the wire tips.

Any refractory material that does not decompose or vaporize can be used for melt spraying. Particles do not coalesce within the spray. The temperature of the particles and the extent to which they melt depend on the flame temperature, which can be controlled by the fuel:oxidizer ratio or electrical input, gas flow rate, residence time of the particle in the heat zone, the particle-size distribution of the powders, and the melting point and thermal conductivity of the particle. Quenching rates are very high, and the time required for the molten particle to solidify after impingement is typically 10⁻⁴ to 10⁻² s.

A broad range of materials can be handled by plasma spraying. However, each material requires some optimization of conditions, such as modification of carrier gas, power, particle-size distribution, and substrate. In some melt-spray processes, a preferential loss of material may occur that leads to a change in stoichiometry. It can be compensated for by adding an appropriate additional gas to the gas stream. Thus, additions of oxygen can keep the reduction of TiO₂ and HfO₂ within acceptable limits, whereas methane reduces the loss of carbon from carbides. A wide selection of coating materials is available. Except for particle-velocity limitations, spray conditions are excellent because of the inert, clean carrier gases used. Inert atmospheres reduce the degradation or oxidation of the particles, preclude oxidation of the substrate, and diminish contamination of the sprayed material resulting from gas adsorption.

Arc wire, flame spraying, HVOF, and arc plasma spraying are the least expensive. Flame plating is more expensive. The adherence of melt-sprayed ceramic coatings onto metallic substrates varies widely from process to process and depends on procedure and substrate preparation (Table 1). The high strength of D-gun and HVOF coatings is a consequence of the higher particle velocity that results in penetration of the substrate surface. For other flame-spray processes, the substrate is usually roughened to provide a suitable surface. Chemical bonding contributes to adhesion but the mechanical bond is preferred. Advantages and disadvantages of flame-spray processes having porosities of 0.1–20.0% are as follow:

Advantages	Disadvantages
produces coatings on any substrate	
versatility in configuration size, shape, and structure	undercuts, deep blind holes, and small internal diameters are difficult to coat
good dimensional control, no shrinkage	
mixed compositions	
fine grain size	properties are generally lower than wrought or pressed and sintered materials
permits patching and repairing	nonstoichiometry may result
alternating layers	
graded coatings	
by masking, localized areas can be coated	residual stresses can limit coating thickness

Table 1. Tensile-Bond Test for Melt-Sprayed Al₂O₃

Method	Particle velocity, m/s	Bond strength, MPa ^a
flame spraying	46–83	3.4–13.8
HVOF ^b	700–800	60–80
arc-wire	15–30	30–70
high power dc	100–1000	15–80
flame plating	700–800	69
d-c plasma process	30–213	6.9–69
induction-plasma process	6–>300	3–15

^a To convert MPa to psi, multiply by 145.

^b HVOF = high velocity oxy fuel, particle velocity 700–800 m/s

Electrophoretic Processes. Electrophoretic coatings are obtained through the migration of charged particles when a potential is applied to electrodes immersed in a suitable suspension of the particles. This process is particularly suited for applying uniform layers on complex bodies. Projecting edges become insulated and the current shifts to bare surfaces as the deposit builds up on them. Particles, not ions, are deposited. The early stages of coating are powdery and densification is required to produce adherence. Electrophoresis can be used to apply metals and alloys, ceramics, and cermets. The three steps are preparation of the dispersion, deposition and conversion, and bonding. In the preparation phase, particle size is critical and diameters may be from 1 to 50 μm; however, too small a size leads to cracking after thick coatings are dried and consolidated (see ELECTROSEPARATION, ELECTROPHORESIS).

Several suspension media having suitable viscosities, densities, conductivities, dielectric constants, and chemical stabilities are available. Binders improve the green strength of the coating. Deposition is effected rapidly using potentials of 50–1000 V dc. Simple d-c sources having no special filtering are sufficient. All inorganic materials are amenable to deposition and most are amenable to codeposition, even though they have different densities; for example, NiO (7.5 g cm³) and WC (15.7 g cm³) have been codeposited in the same ratio as they were dispersed. The practical limit of thickness is about 0.5 mm as set by shrinkage cracking. Conversion and bonding techniques vary with the type of coating. Hydrogen reduction, hydrostatic pressing, and sintering are used for most metal-base coatings. The oxides of iron, nickel, cobalt, molybdenum, and tungsten can be reduced in hydrogen at less than 600°C. For example, NiO–Cr green deposits are reduced in H₂ at 320°C to produce metallic nickel, with no particle sintering. After pressing at 69 MPa (10,000 psi), further densification is achieved by sintering to 90% of theoretical density at 1090°C. Vacuum sintering is preferred for titanium, zirconium, niobium, and tantalum. Glass and ceramic coatings can be either vitrified or fused.

Bulk Coating

In bulk coating processes, bulk materials are joined to the substrate either by a surface melt process or by attachment of the solid material. An example of the latter is the application of heat-resistant tiles of silica-type material to the aluminum alloy skin of a space shuttle vehicle, enabling the vehicle to withstand the reentry heat.

In cladding, one metal is coated with another by rolling or extruding the two metals in close contact to each other. Coherence of the two metals is induced by soldering, welding (qv), or casting one in contact with the other prior to the rolling operation. By this mechanical process, steel can be coated with copper, nickel, or aluminum. The coextrusion process, which is used for lightweight rifle barrels made of a steel core and titanium alloy wrap, is a good example of the cladding process.

Another familiar commercial method is the immersion or hot-dipping process. The article to be coated is immersed in a molten metal bath. Usually little else is done to change the properties of the coating, which adheres to the surface upon removal of the article from the bath. For a successful coating, an alloying action must take place between the components to some extent. Zinc and tin coatings are applied to sheet steel by hot-dipping.

Surface coatings are also applied by welding processes, such as manual arc welding (oxy–acetylene, gas–tungsten), gas–metal arc welding, submerged-arc welding, spray welding, plasma-arc welding, and electroslag welding (see WELDING). Thin coatings (0.25–0.75 mm) are applied by spray welding. For heavier coatings in the range of 6 to 65 mm or more, other welding processes are used.

These welding processes are not suitable for the application of refractory coatings of reactive metals such as tungsten, tantalum, niobium, and chromium, although electroslag welding is being studied for this purpose. On the other hand, such refractory metal, alloy, and cemet coatings can be applied by electron beam melting to form thick coatings, because this process is carried out in vacuum.

A new cladding process permits brazing of carbides, nitrides, and metals and alloys to most metallic substrates. These thick coatings contain 95% metal (<5% braze component) and are metallurgically attached to the substrate. Another development in this area is the use of high power laser beams to surface melt refractory metal or cemet coatings onto substrates in a controlled atmosphere to prevent contamination of the reactive metals. This process is sometimes referred to as laser glazing (see LASERS). Significant surface stresses are produced but pore-free barriers often result.

Reaction spraying is also being widely explored. In this case new compositions or particle surface reactions providing enhanced coating properties, eg, nitrides for wear resistance, are achieved while the particle is in transit.

Enameling meets decorative as well as protective requirements. Ceramic enamels are mainly based on alkali borosilicate glasses. The part to be enameled is dipped into or sprayed with a slip, ie, a water suspension of glass fragments called frit. The slip coating is dried and fused in an enameling furnace under careful heat control (see ENAMELS, PORCELAIN OR VITREOUS).

Troweling and painting methods are used to apply thick protective coatings of a refractory paste or cemet onto a variety of substrates for high temperature service. Fiberfrax (The Carborundum Company) coating cements, composed of aluminosilicate fibers bonded with air-setting temperature-resistant inert binders, is commonly used as a coating for reducing the oxidation rate of graphite.

In the microelectronics industry, powdered metals and insulating materials that consist of nonnoble metals and oxides are deposited by screen printing in order to form coatings with high resistivities and low temperature coefficients of resistance. This technique may be useful in depositing oxide–metal refractory coatings.

Surface-Modification Processes

Cementation or Diffusion Coatings. Cementation is defined as the introduction of one or more elements into the outer portion of a metal object by means of diffusion at elevated temperatures. Cementation was first used to convert iron to steel, and copper to brass, but as of the mid-1990s it is considered only as a method of surface treatment. The coating produced by cementation is formed by an alloying or chemical combination of the diffusing elements and the substrate material. Cementation coatings enjoy wide metallurgical application. A prime example is the case hardening of steel whereby a soft, ductile, low carbon steel is heated to 810–900°C in a packing of carbonaceous material to produce a high carbon steel surface (see METAL SURFACE TREATMENTS; STEEL). The coating formed in this manner is limited to use at temperatures of about 600°C or lower. Truly refractory coatings are alloyed onto molybdenum, tantalum, niobium, and tungsten by cementation. Such coatings provide short-term protection at 1650°C or higher, and long-term protection at 1370°C.

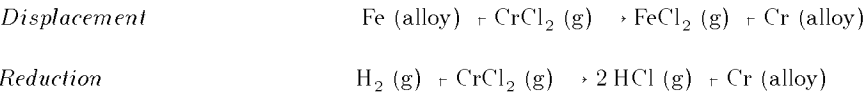
Cementation coatings are produced at temperatures well below the melting points of either the coating or substrate material by means of a vapor-transport mechanism. The coating material and the substrate form definite chemical compounds, such as silicides, aluminides, beryllides, or chromides. On the other hand, some cementation coatings can be solid solutions of indefinite composition. Oxidation-resistant coatings act as diffusion barriers for both the inward diffusion of oxygen and the outward diffusion of the substrate. The diffusion of the coating components into the substrate is generally the rate-controlling step. Cementation coatings are applied by pack cementation, activated or nonactivated slurry processes, and the fluidized-bed technique.

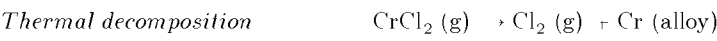
In pack cementation, the part to be coated is placed in a retort and surrounded with a powdered pack consisting of the coating component and an activator; the latter reacts with the coating component to form the carrier vapor, usually a halide or an inert diluent, to prevent the pack from sintering together and to permit vapor transport of the alloying component through the pack.

The slurry process requires less coating component. The latter is suspended in a vehicle, eg, lacquer or water, and is painted onto the substrate. The coated part is heated in an alumina retort containing a layer of activator at the bottom. The coating component forms a halide and is deposited onto and diffused into the substrate. Slurry processes can be either activated or nonactivated. In the latter case, development of the coating relies purely on diffusion without the possible benefits of vapor deposition.

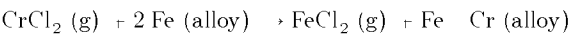
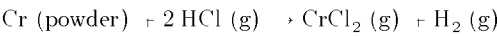
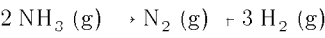
The fluidized-bed technique combines aspects of pack cementation and vapor deposition. A fluidized bed consists of a mass of finely divided solids contained in a column. The solids are brought into a fluidized state by the lifting action of a gas as it rises through the column. A vaporized halide may be carried by a gas into the bed, where it reacts with the fluidized coating powder to form the coating component halide, which then thermally decomposes and deposits on the substrate contained within the retort. Alternatively, the metal coating particles may be fluidized before entering the bed, whereby the halide gas permeates the fluidized particles to form the coating halide vapor. This vapor is carried into the bed of inert particles, where it thermally decomposes to deposit the coating (see FLUIDIZATION).

The chromizing of iron is described by three mechanisms:



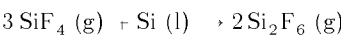
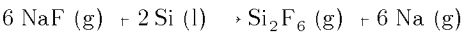


The displacement mechanism involves placing the iron alloy packed in chromium powder, NH_4Cl , and Al_2O_3 in a sealed retort, which is heated to promote vapor deposition and diffusion processes. The exact chemistry is not known, but the following steps probably occur:



Several thermodynamic and kinetic requirements must be fulfilled for the reactions to proceed as shown. First, the vapor pressure of the activators or carriers must be sufficiently high at the coating temperature for the reaction to proceed. The activator, ie, the chloride, should have a boiling point slightly above the temperature of the coating to provide a reservoir for reaction without volatilizing too rapidly. The coating-component halide should have a boiling point below the coating temperature to saturate the pack. The volatility of the by-product of the coating reaction must be high; fast removal prevents the formation of a barrier to continued deposition of the coating element and also prevents contamination of the coating.

Another example is the siliciding of tantalum, basically an oxidation–reduction reaction. The packing is sodium fluoride and silicon. After deposition, the coating diffuses continuously into the substrate, according to the following reactions:



If the rules for volatiles and thermodynamics of the halides are followed, the reaction can be used for aluminizing, siliciding, chromizing, and similar processing.

Silicide coatings of refractory metals may contain as much as three to five coating components other than silicon. A mixture of halide carriers is selected containing the best carrier for each component.

The outstanding characteristics of a fluidized bed are its high heat-transfer coefficient and its turbulence, which yield optimum temperature uniformity throughout the bed. These factors contribute to the successful treatment of large, complex objects, which might not be possible by other means.

Analogous to the carburization of steel is the nitrogen case-hardening of steel in which surfaces are hardened by heating the steel in the presence of nascent nitrogen. The nitrogen reacts with impurities or alloy constituents, eg, aluminum, chromium, vanadium, tungsten, etc, to form dispersed nitrides. Other examples of cementation coatings include the surface-hardening of steel by immersion in a molten sodium cyanide bath at 850–870°C to promote codiffusion of both carbon and nitrogen into the surface. Iron and steel are cemented with silicon at 800°C, and iron, nickel, and copper with tungsten at 800–1350°C. Similarly, iron and nickel are cemented with molybdenum by heating in ferromolybdenum. In the same temperature range, titanium coatings are obtained on ferrous metals by heating them for 1.5 h in a powdered mixture of sponge titanium containing 0.5–6.0% Fe at 800–1200°C. Boride coatings are produced by heating iron, cobalt, or nickel substrates in boron powder at 950°C in a vacuum of 67 mPa (5×10^{-4} mm Hg).

The most advanced cementation coatings are the intermetallic coatings, specifically silicides and aluminides, that protect refractory metals. The earliest and simplest of these is the molybdenum silicide coating developed for molybdenum substrates. It consists largely of MoSi_2 , but is modified by additions of boron, manganese, titanium, chromium, beryllium, etc, singly or in combinations. Coating systems for niobium are more complex, but oxidation-resistant coatings result by diffusing silicon, chromium, aluminum, boron, titanium, and beryllium singly or in combinations into the substrate. Diffusion coatings for tantalum are based on the formation of TaAl_3 , an oxidation-resistant material. Silicide coatings on tantalum also provide significant oxidation protection. Aluminum, boron, or manganese are often added as modifiers. Tungsten presents a more difficult problem, but various silicide coatings provide a measure of protection at temperatures as high as 1815°C. After 10 h at this temperature, WSi_3 is converted to the glassy $\text{W}_5\text{Si}_4\text{O}_2$ composition.

An activated slurry process can be used to place impervious silicon carbide coatings on graphite. A slurry of silicon carbide, carbon, and appropriate organic binders is applied to the surface of the graphite by spraying, dipping, or painting. After a low temperature treatment to drive off the binder, the coated substance is heated in the presence of silicon vapor, which diffuses into the surface to form a SiC layer on the graphite. The coating composition can vary from self-bonded silicon carbide containing only a few percent of uncombined silicon to SiC crystals bonded with a continuous silicon matrix. Successful SiC coatings on complex graphite shapes have been obtained with isotropic graphite substrates possessing compatible thermal expansion coefficients. Resistance to oxidation at 1400°C for 100 h or more has been accompanied by successful service in nuclear reactor environments. Coatings of SiC on graphite fail under compressive loads or impact at levels that do not damage uncoated graphite. This occurs because the graphite is less brittle and deforms under load, whereas the thin, more rigid SiC coating does not.

Aluminide and silicide cementation coatings such as TaAl_3 on tantalum and MoSi_2 on molybdenum oxidize at slow rates and possess some inherent self-repair characteristics. Fine cracks that appear and are common to these coatings can be tolerated because stable, protective oxides form within the cracks and seal them. Thermal cycling, however, accelerates failure because of thermal expansion mismatch that ultimately disrupts the protective oxide coating.

An important application is the aluminizing of air foils of gas-turbine engines made of high temperature Ni- or Co-base alloys. The aluminizing can be carried out either in a pack process or in an out-of-the-pack process.

Cementation coatings rely on diffusion to develop the desired surface alloy layer. Not only does the coating continue to diffuse into the substrate during service, thereby depleting the surface coating, but often the substrate material diffuses into the surface where it can be oxidized. Because the diffusion rate is temperature dependent, this may occur slowly at lower service temperatures.

The substrate has to be prepared for cementation. The surface must be clean and free of oxide. Corners and edges are particularly important in diffusion-type coatings; sharp edges are usually detrimental. Barrel finishing, ie, tumbling in a barrel with abrasive media, may result in the desired shape.

The quality of cementation coatings does not necessarily equal that produced by other techniques. There are cases, however, where a moderate degree of corrosion resistance is useful and where other requirements are best met by the application of cementation. Using other processes, it may be difficult to coat porous surfaces or preserve the contour of machined surfaces. Cementation may then be the method of choice. Coatings for molybdenum, niobium, tantalum, and tungsten not easily obtained by other means can be achieved by this method. Coated refractory metal systems have been applied to leading edges, skins, and structural members of space vehicles, rocket combustion chambers, nozzle inserts, extension skirts, and to vanes or blades for advanced gas-turbine engines.

Metalliding. Metalliding, a General Electric Company process (9), is a high temperature electrolytic technique in which an anode and a cathode are suspended in a molten fluoride salt bath. As a direct current is passed from the anode to the cathode, the anode material diffuses into the surface of the cathode, which produces a uniform, pore-free alloy rather than the typical plate usually associated with electrolytic processes. The process is called metalliding because it encompasses the interaction, mostly in the solid state, of many metals and metalloids ranging from beryllium to uranium. It is operated at 500–1200°C in an inert atmosphere and a metal vessel; the coulombic yields are usually quantitative, and processing times are short; controlled

uniform coatings from a few to many μm are obtained, many of which are unavailable by other techniques. Diffusion rates are high and the process can be run continuously. Boron and silicon anodes can be diffused into most metals of Groups 5–11 (VB, VIB, VIIB, VIII, and IB) of the Periodic Table.

The borides are extremely hard (9.8–29 GPa (1000–3000 kgf mm^2) Knoop) and, in the case of molybdenum, >39 GPa (4000 kfg mm^2) (see **HARDNESS**). However, oxidation resistance is usually poor unless a subsequent coating is formed, such as silicidizing or chromizing, which imparts oxidation resistance. Silicides are generally very oxidation resistant, but not as hard as borides. Silicide coatings formed on molybdenum (51 μm in 3 h) at 675°C have superior oxidation resistance. At these low temperatures, the molybdenum substrate does not embrittle and the coatings are quite flexible.

The metals that can be aluminized act similarly with boron, as do the metals in Group 4 (IVB). Resistance to oxidation is the principal benefit. Titanizing and zirconizing are extremely sensitive to oxygen impurities, and when the salts are completely free of oxides and blanketed by high purity argon, excellent diffusion coatings can be formed in LiF at 900–1100°C. The most promising area for applications appears to be with nickel- and iron-based alloys where intermetallic compound formation gives rise to many unique coatings that are tough and oxidation resistant. Beryllide coatings can be formed on approximately 40 metals ranging from titanium to uranium and with compositions such as TiBe_{12} , $\text{Ni}_5\text{Be}_{21}$, and UBe_{13} . They are very hard, usually oxidation resistant, and easily formed up to several mils in thickness.

Moving further to the right in the Periodic Table, the scope of the metallizing processes becomes much more limited. Iron, cobalt, and nickel are restricted to approximately 12 metals into which they can be diffused. Iron can be diffused into cobalt and nickel with very good results, but the reverse is unsuccessful. The diffusion of nickel into molybdenum, tungsten, and copper has been successful. Germanium occupies a slightly higher position in the electromotive series than nickel and therefore diffuses into nickel; it is below cobalt and iron, however, as determined by voltage measurements in the salts, and cannot be diffused into either of these metals, even at high voltages and current densities. It is readily diffused into molybdenum, palladium, copper, platinum, gold, and Monel.

Microstructure of Coatings

The microstructure of bulk coatings resembles the normal microstructure of metals and alloys produced by melt solidification. The microstructure of particulate-deposited materials resembles a cross between rapidly solidified bulk materials having severe deformation and powder compacts produced by pressing and sintering. A special feature of particulate coatings is a significant degree (ca 2–20 vol %) of porosity that strongly affects the properties of the deposit.

The microstructure and imperfection content of coatings produced by atomistic deposition processes can be varied over a very wide range to produce structures and properties similar to or totally different from bulk processed materials. In the latter case, the deposited materials may have high intrinsic stress, high point-defect concentration, extremely fine grain size, oriented microstructure, metastable phases, incorporated impurities, and macro- and microporosity. All of these may affect the physical, chemical, and mechanical properties of the coating.

PVD Condensates. Physical vapor deposition condensates can deposit as single-crystal films on certain crystal planes of single-crystal substrates, ie, by epitaxial growth or, in the more general case, the deposits are polycrystalline. In the case of films deposited by evaporation techniques, the main variables are the nature of the substrates; the temperature of the substrate during deposition; the rate of deposition; and the deposit thickness. Contrary to what might be expected, the deposit does not initially form a continuous film of one monolayer and grow. Instead, three-dimensional nuclei are formed on favored sites on the substrates, such as cleavage steps on a single-crystal substrate. These nuclei grow laterally and in thickness (growth state), ultimately impinging on each other to form a continuous film. The average thickness at which a continuous film forms depends on the nucleation density and the deposition temperature and rate; both influence the surface mobility of the adatom. This thickness varies from 1 nm for Ni condensed at 15 K to 100 nm for Au condensed at 600 K.

The microstructure and morphology of thick single-phase films have been extensively studied for a wide variety of metals, alloys, and refractory compounds. The structure model first proposed is shown in Figure 6 (10). It was subsequently modified as shown in Figure 7 (10,11).

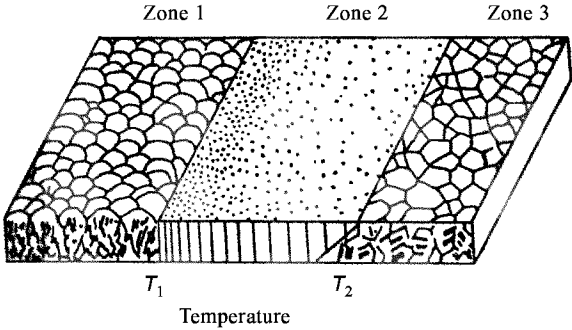


Fig. 6. Structural zones in condensates at various substrate temperatures. When T_m is the melting point (10):

	Zone 1	Zone 2	Zone 3
metals	$<0.3 T_m$	$0.3\text{--}0.45 T_m$	$>0.45 T_m$
oxides	$<0.26 T_m$	$0.26\text{--}0.45 T_m$	$>0.45 T_m$

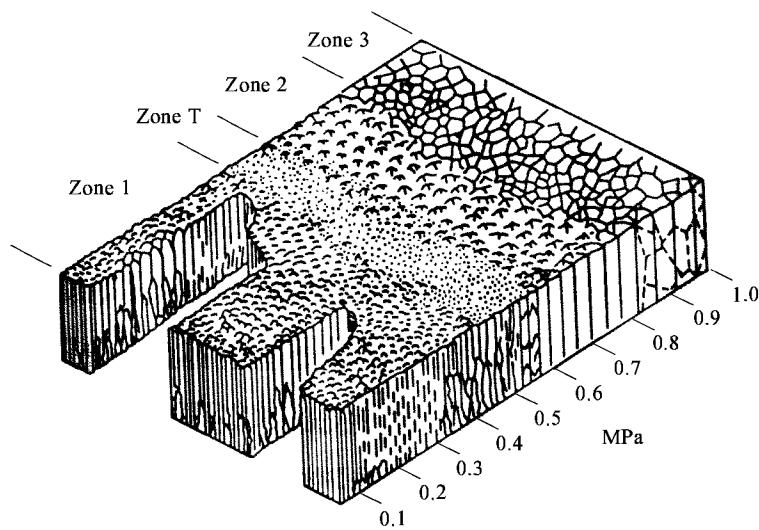


Fig. 7. Structural zones in condensates showing the effect of gas pressure. To convert MPa to psi, multiply by 145. See text.

At low temperatures, the surface mobility of the atoms is limited and the structure grows as tapered crystallites from a limited number of nuclei. It is not a full density structure but contains longitudinal porosity on the order of a few tens of nm width between the tapered crystallites. It also contains numerous dislocations with a high level of residual stress. Such a structure has also been called botryoidal and corresponds to Zone 1 in Figures 6 and 7.

As the substrate temperature increases, the surface mobility increases and the structural morphology first transforms to that of Zone T, ie, tightly packed fibrous grains having weak grain boundaries, and then to a full density columnar morphology corresponding to Zone 2 (see Fig. 7).

The size of the columnar grains increases as the condensation temperature increases. Finally, at still higher temperatures, the structure shows an equiaxed grain morphology, Zone 3. For pure metals and single-phase alloys, T_1 is the transition temperature between Zone 1 and Zone 2 and T_2 is the transition temperature between Zone 2 and Zone 3. According to the original model (10), T_1 is $0.3 T_m$ for metals and $0.22\text{--}0.26 T_m$ for oxides, whereas T_2 is $0.4\text{--}0.45 T_m$ for both (T_m is the melting point in K) (see Fig. 6).

The modification shows that the transition temperature may vary significantly from those stated above and in general shift to higher temperatures as the gas pressure in the synthesis process increases. The transition from one zone to the next is not abrupt, but smooth. Hence, the transition temperatures should not be considered as absolute but as guidelines. Furthermore, not all zones are found in all types of deposit. For example, Zone T (see Fig. 7) is not prominent in pure metals, but becomes more pronounced in complex alloys, compounds, or in deposits produced at higher gas pressures. Zone 3 is not often seen in materials with high melting points.

CVD Coatings. As in PVD, the structure of the deposited material depends on the temperature and supersaturation, roughly as pictured in Figure 8 (12). In the case of CVD, however, the effective supersaturation, ie, the local effective concentration in the gas phase of the materials to be deposited, relative to its equilibrium concentration, depends not only on concentration, but on temperature. The reaction is thermally activated. Because the effective supersaturation for thermally activated reactions increases with temperature, the opposing tendencies can lead in some cases to a reversal of the sequence of crystalline forms listed in Figure 8, as temperature is increased (12).

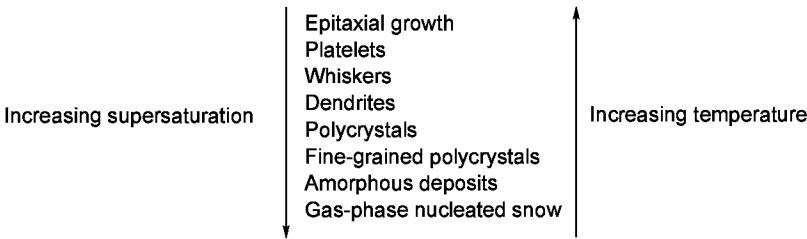


Fig. 8. Morphological effects of supersaturations and temperature on vapor deposited materials (12).

Growth of columnar grains is characteristic of many materials in certain ranges of conditions. This structure results from uninterrupted growth toward the source of supply. Where growth in one crystallographic direction is preferred over others, grains having that orientation engulf those of other orientations.

Electrodeposits. Columnar structures are characteristic of deposits from solutions, especially acid solutions, containing no additives, high metal-ion concentration solutions with high deposition rates, or from low metal-ion concentration solutions at low deposition rates. These usually exhibit lower tensile strength, elongation, and hardness than other structures, but are generally more ductile. Such deposits are usually of highest purity (high density) and low electrical resistivity.

Fibrous structures represent a grain refinement of columnar structure. Stress-relieving additives, eg, saccharin or coumarin, promote such refinement, as do high deposition rates. These may be considered intermediate in properties between columnar and fine-grained structures.

Fine-grained deposits are usually obtained from complex ion solutions, eg, cyanide, or using certain addition agents. These deposits are less pure, less dense, and exhibit higher electrical resistivities because of the presence of foreign material.

Banded structures are characteristic of some alloy deposits and of bright deposits resulting from brightening addition agents. Plating-current modifications (periodic reverse (PR), interrupted current (IC), pulse) favor the conversion of normal structure from a solution to a banded structure. These deposits generally possess higher tensile strength, hardness, and internal stress and lower ductility than the other structures.

Grain size varies widely, from 10 to 5000 nm. The grain size of fine-grained or banded deposits is usually 10–100 nm. Some metals, notably copper, nickel, cobalt and gold, can be deposited in all four types of grain structure, depending on the solution composition and plating conditions.

Characterization and Testing

Evaluation for high temperature service has been less reliable and less standardized than evaluations for conventional service. This is partly the result of the difficulty of reproducing in the laboratory the severe service conditions capable of yielding acceptable correlations. Test conditions do not necessarily simulate the geometrical and environmental conditions of the service. Nevertheless, screening tests yield primary behavioral parameters that usually define the limits of operation. Various tests are employed for the characterization of refractory coatings: optical and electron microscopy and metallographic and microscopic observation, which evaluate substrate and coating structures, coating thickness, bond characteristics, and detection of inclusions; electron microscope and scanning and transmission electron microscopy, which evaluate detection of injurious inclusions in the substrate or coating; x-ray

diffraction (electron), which evaluates the effects of processing on coating composition, ie, composition of coating at various depths within the coating can be determined by controlled polishing followed by x-ray; chemical microscopy, eg, Auger electron spectroscopy, electron spectroscopy for chemical analysis, and secondary ion mass spectroscopy, which determine chemical analysis of surface and subsurface layers, ie, resolution can be as low as 5-nm segregation at imperfections; bend, tensile, and tensile bond tests, which determine adhesion, the ductility effect of coating or processing on base-material strength, ductility, and elongation; microhardness transverse and corrosion tests, which gauge the cross-sectional hardness of the coating and substrate, the effect of processing on the substrate, and ductility; fatigue tests, which show the effect of coating or coating process on the substrate, fatigue properties, and fatigue strength of a coated-metal system; thermal tests, eg, chemical oxidation and thermal cycling with sustained load in air, which evaluate oxidation and thermal shock resistance, the effect of coating on ductility of base and specimen creep on coating, integrity, and thermal shock and oxidation resistance; and the plasma or oxyacetylene torch test, which shows coating emissivity, thermal shock and oxidation resistance, and melting point.

Nondestructive tests for refractory coatings include the following: evaporgraph which detects residual moisture left behind in cracks, pores, or flaws, and evaluates discontinuities on flat panel surfaces; electromagnetic inspection, which evaluates pores, cracks, and pits; ultrasonic inspection, which shows surface and subsurface flaws, unsuitable for thin skins; fluorescent particle inspection, ie, finely divided, fluorescent-coated magnetic particles attracted to and outlining the pattern of any magnetic leakage fields created by discontinuities, which determine surface cracks, pits, and similar coating defects; red-dye penetrant, which shows surface flaws; radiographic inspection, which shows small coating flaws in assembled structures, although the sensitivity is difficult to control; and microscopy, which evaluates flaws by observation of oxidation and weakness in bond strength resulting from thermal shock. This test is very reliable after an exposure test. Many of these factors are unsuitable for corner or edge defects.

Selection Criteria

The selection of a particular deposition process depends on the material to be deposited and its availability; rate of deposition; limitations imposed by the substrate, eg, maximum deposition temperature; adhesion of deposit to substrate; throwing power; apparatus required; cost; and ecological considerations. Criteria for CVD, electrodeposition, and thermal spraying are given in Table 2 (13).

Table 2. Characteristics of Deposition Processes

Characteristic	Evaporation	Ion plating	Sputtering	Chemical vapor deposition	Electro-depositio n	Thermal spraying
mechanism of production of depositing species	thermal energy	thermal energy	momentum transfer	chemical reaction	deposition from solution	from flames, arcs, or plasmas
deposition rate, nm min	<75,000	<25,000	low except for pure metals ^a	moderate, eg, 20–2500	low to high	very high
depositing specie	atoms and ions	atoms and ions	atoms and ions	atoms	ions	droplets
throwing power for complex shaped object	poor line-of-sight coverage ^b	good ^c	good ^c	good	good	none
into small blind holes	poor	poor	poor	limited	limited	very limited
deposition of metal	positive	positive	positive	positive	positive but limited	positive
alloy refractory compound	positive	positive	positive	limited	limited	positive
energy, depositing species, eV	ca 0.1–0.5	1–100	1–100	positive high for PACVD	limited	positive
bombardment of substrate and deposit by inert gas ions	normally not	yes	possible, depending on geometry	possible	can be high	positive
growth interface perturbation	normally not	yes	yes	yes, by rubbing	none	none
substrate heating by external means	normally yes	yes	generally not	no	none	normally not

^a For copper, 1000 nm min.
^b Except by gas scattering.
^c Thickness distributions may be nonuniform.

Applications

Coatings can be classified into six categories: chemically functional, mechanically functional, optically functional, electrically functional, biomedical, and decorative. In addition, there are some unique applications in the aerospace program, such as the ablative coatings of pyrolytic carbon and graphite- and silica-based materials for protection of nose cones and the space shuttle during reentry (see ABLATIVE MATERIALS). Another unique energy-related application is the coating of low atomic number (low-Z) elements such as TiC for the first wall of thermonuclear reactors to minimize contamination of the plasma.

Chemically Functional. Refractory coatings are used for corrosion-resistant high temperature service in gas turbine and diesel engines, components such as crucibles, thermocouple protection tubing, valve parts, etc.

Blades and vanes used in the hot-end of a gas turbine are subject to high stresses in a highly corrosive environment of oxygen, sulfur, and chlorine-containing gases. A single or monolithic material such as a high temperature alloy cannot provide protection against both. A bulk alloy designed for its mechanical properties provides the corrosion resistance by means of an overlay coating of an M–Cr–Al–Y alloy where M stands for Ni, Co, or Fe or Ni + Co. In production, the coating is deposited by electron beam evaporation; in the laboratory, by sputtering or plasma spraying. These overlay coatings have several advantages over diffusion aluminide coatings. The latter lose their effectiveness at higher temperatures because of interactions with the substrate. The composition and properties of overlay coatings can be more easily tailored to the needs of specific applications. In addition, coatings of stabilized zirconia are used as thermal barriers in diesel engines and gas turbines to raise operating temperature of the engine and protect it from corrosive fuels.

Boron nitride and titanium diboride coatings are used on graphite for the evaporation of liquid aluminum.

Mechanically Functional. Refractory coatings are used in engine parts, landing gears, soft-film lubricants, and cutting and forming tools (see TOOL MATERIALS).

A large and rapidly growing application is the coating of cutting and forming tools and industrial knives with carbides, nitrides, oxides, or multiple layers of these materials. These coatings are deposited by CVD and PVD methods and increase the tool life by factors ranging from 2 to 10, depending on

the operating conditions. Cermet coatings such as TiB₂–Ni are also deposited electrolytically to provide wear resistance.

These coatings are also employed as solid lubricants in engine components. Refractory materials such as MoS₂ and WSe₂ are lamellar compounds and provide very effective solid-state lubrication in spacecraft bearings and components used in radiation environments where conventional organic liquid lubricants are not stable (see LUBRICATION AND LUBRICANTS).

Optically Functional. Laser optics, layer architectural glass panels (up to 3 × 4.3 m), lenses, TV-camera optical elements, and similar applications require optically functional coatings. A large and expanding operation is coatings on architectural glass panels used in buildings to alter the transmission and reflection properties of glass (qv) in various wavelength ranges and thus conserve energy usage. Furthermore, attractive visual effects, such as a bronze appearance, can be achieved. These coatings consist of multiple layers of oxides and metals (the precise compositions are proprietary) and are deposited in large in-line sputtering and evaporation systems. Future applications should utilize the selective transmission of coatings of nitrides, carbides, and borides in solar-thermal applications.

Another growing application that overlaps the electrically functional area is the use of transparent conductive coatings or tin oxide, indium–tin oxide, and similar materials in photovoltaic solar cells and various optic electronic applications (see PHOTOVOLTAGIC CELLS). These coatings are deposited by PVD techniques as well as by spray pyrolysis, which is a CVD process.

Electrically Functional. Refractory coatings are used in semiconductor devices, capacitors, resistors, magnetic tape, disk memories, superconductors, solar cells, and diffusion barriers to impurity contamination from the substrate to the active layer.

Thin-film capacitors and resistors contain such dielectric materials as silicon monoxide and dioxide, tantalum oxide, silicon nitride, and the like. These coatings are deposited by a variety of atomic deposition techniques (PVD, CVD) and thick-film methods. In some cases coatings, eg, SiO₂, are formed by thermal oxidation. An important application is the deposition of passivating layers of SiO₂ or SiO₂–Si₃N₄ by plasma-assisted CVD techniques. A newer area is insulation coatings for GaAs devices formed by plasma-assisted oxidation methods and the deposition of amorphous silicon by CVD techniques. The fabrication of tungsten-emitter elements for thermionic convertors by CVD is another application. Thermal sprayed coatings are used widely for thicker spot coatings (wire attachment) and current conductors, ie, on collectors (silicon solar cells). Typical products in use include polypropylene foil capacitors, material components, varistors, etc.

Biomedical. Heart-valve parts are fabricated from pyrolytic carbon, which is compatible with living tissue. Such parts are produced by high temperature pyrolysis of gases such as methane. Other potential biomedical applications are dental implants and other prostheses where a seal between the implant and the living biological surface is essential. Plasma and arc-wire sprayed coatings are used on prosthetic devices, eg, hip implants, to achieve better bone tissue attachments (see PROSTHETIC AND BIOMEDICAL DEVICES).

Decorative. Titanium nitride has a golden color and is used extensively to coat steel and cemented carbide substrates for watch cases, watch bands, eyeglass frames, etc. It provides excellent scratch resistance as well as the desired aesthetic appearance, and it replaces gold coatings used previously.

Economic Aspects

Diffusion aluminide and silicide coatings on external and internal surfaces for high temperature corrosion protection in parts such as gas-turbine blades is estimated at \$40 × 10⁶ yr in North America and about \$50 × 10⁶ worldwide.

Overlay coatings onto gas-turbine blades and vanes of M–Cr–Al–Y type alloys by electron beam evaporation is estimated at \$10 × 10⁶ to coat 200,000 parts at an average cost of \$50 per part.

Hard facing of various components in the aircraft gas-turbine engine and in industrial applications for textile machinery parts, oil and gas machinery parts, paper-slitting knives, etc, is estimated at \$1 × 10⁹ in 1995 with an estimated growth rate of 5° annually. The mix is approximately 45° aerospace applications, 55° industrial applications. Additionally, repair coatings for gas-turbine blades and vanes is estimated at \$500 × 10⁶. These coatings are primarily deposited by plasma spray, arc-wire, HVOF, and detonation gun techniques.

Refractory compound coatings of carbides, nitrides, and oxides on cemented carbide cutting tools, mainly by the CVD process, are estimated at \$300 × 10⁶ annually worldwide.

Another application is the coating of complex-shaped high speed steel cutting tools such as drills, holes, gear-cutters, etc, using a titanium nitride coating deposited by the methods of plasma-aided reactive evaporation and ion plating. A very rough estimate of the add-on value of the coating is \$(50) × 10⁶ in 1995, but with a rapid growth rate of 100–300° expected into the year 2000. The same processes are also used to deposit gold-colored wear-resistant titanium nitride coatings on watch bezels, watchbands, and other decorative jewelry items. A very rough estimate is \$20 × 10⁶ annually. The last two applications are practiced principally in Japan, Europe, and the United States.

Refractory coatings extensively used in the semiconductor and optical materials industry have an annual add-on value exceeding \$10⁹.

The thermal spraying industry had rapid, 30° annual growth from 1960–1989 because of the acceptance of this process by the aircraft gas-turbine industry. Over 300 separate coating areas were used in each engine as of 1995. This application area was 50° of the total thermal spraying market. A slump in this market niche reduced industry sales by 30° from 1990 to 1995. Growth has resumed owing to broader acceptance by other industry segments, such as the automotive industry. Significant merger activity in the early 1990s reduced ownership in this industry, formerly dominated by U.S. companies, to two principal players: Sulzer Metco (Wohlen, Switzerland), which had annual sales in 1995 of \$120 × 10⁶, and Eutetic Castolin (Lausanne, Switzerland), with \$80 × 10⁶. In contrast, Praxair Materials (Indiana) had only \$60 yr.

There are job shops and applicators in the thermal spraying business and the annual value was about \$1 × 10⁹ in 1995. Up to 65° of the total company sales represented materials prices from \$1–\$25 kg. Equipment prices for spray technology and the number of units sold per annum in the mid-1990s are given in Table 3.

Table 3. Spraying Processes Equipment Prices

Equipment type	Price, \$ × 10 ³	Units sold yr
<i>Spray equipment</i>		
plasma	25–175	100
wire spray	10–30	1000
^a	40–90	250
large combustion powder	5–50	300
<i>Automation equipment</i>		
automated cells	100–2000	50
water jet coating strippers	100–1000	30

^a HVOF high velocity oxy fuel.

BIBLIOGRAPHY

"Refractory Coatings" in *ECT* 2nd ed., Vol. 17, pp. 217–284, by B. R. Miccioli, The Carborundum Co.; in *ECT* 3rd ed., Vol. 20, pp. 38–64, by R. F. Bunshah, UCLA.

1. H. H. Hausner, ed., *Coatings of High Temperature Materials*, Plenum Publishing Corp., New York, 1966.
2. R. F. Bunshah and D. M. Mattox, *Phys. Today* **33**, 50 (May 1980).
3. R. F. Bunshah, in *New Trends in Materials Processing*, American Society of Metals, Metals Park, Ohio, 1974, pp. 200–269.
4. R. F. Bunshah and A. C. Raghuram, *J. Vac. Sci. Technol.* **9**, 1385 (1972).
5. W. Grossklaus and R. F. Bunshah, *Int. Vac. Sci. Technol.* **13**, 532 (1975).
6. M. Kobayashi and Y. Doi, *Thin Solid Films* **54**, 57 (1978).
7. K. K. Yee, *Int. Metal. Rev.* **1**(226), 19 (1978).
8. T. Bonifield, in R. F. Bunshah, ed., *Films and Coatings for Technology*, Noyes Data Corp., Park Ridge, N.J., 1982, Chapt. 9.
9. **U.S. Pats. 3,024,175–3,024,177, 3,232,853** (Mar. 6, 1962), N. C. Cook (to General Electric Co.).
10. B. A. Movchan and A. V. Demchishin, *Fizika Metall.* **28**, 653 (1969).
11. J. A. Thornton, *J. Vac. Sci. Technol.* **12**, 830 (1975).
12. J. A. Blocher, in Ref. 8, Chapt. 8.
13. R. F. Bunshah and co-workers, in Ref. 8.

General References

Ref. 8 is a general reference.

H. H. Hausner, ed., *Coatings of High Temperature Materials*, Plenum Publishing Corp., New York, 1966.

ASM International, Materials Park, Ohio. Miscellaneous thermal spray conference proceedings.

J. Huminik, Jr., ed., *High-Temperature Inorganic Coatings*, Reinhold Publishing Corp., New York, 1963.

D. H. Leeds, in J. E. Hove and W. C. Riley, eds., *Ceramics for Advanced Technologies*, John Wiley & Sons, Inc., New York, 1965, Chapt. 7.

F. A. Lowenheim, ed., *Modern Electroplating*, 2nd ed., John Wiley & Sons, Inc., New York, 1963.

C. F. Powell, J. H. Oxley, and J. M. Blocher, Jr., *Vapor Deposition*, John Wiley & Sons, Inc., New York, 1966.

R. F. Bunshah, ed., *Techniques of Metals Research*, Vol. I, Parts 1–3; Vol. VII, Part 1, John Wiley & Sons, Inc., New York, 1968.

L. Holland, *Vacuum Deposition of Thin Films*, Chapman & Hall, London, 1968.

L. I. Maissel and R. Glang, *Handbook of Thin Film Technology*, McGraw-Hill Book Co., Inc., New York, 1970.

B. Chapman and J. C. Anderson, eds., *Science and Technology of Surface Coatings*, Academic Press, Inc., New York, 1974.

J. L. Vossen and W. Kern, eds., *Thin Film Processes*, Academic Press, Inc., New York, 1978.

R. F. Bunshah, *Materials Coating Techniques*, Agard Lecture Series, No. 106, NATO, 1980.

A. R. Reinberg, *Annual Rev. Mat. Sci. Technol.* **9**, 341 (1979).

R. W. Haskell and J. G. Byrne, in H. Herman, ed., *Treatise on Materials Science and Technology*, VI, Academic Press, Inc., New York.

R. Bakish, ed., "Electron and Ion Beam Science and Technology," *International Conferences*.

R. Sard, H. Leidheiser, Jr., and F. Ogburn, eds., *Properties of Electrodeposits—Their Measurement and Significance*, Electrochemical Society, Princeton, N.J.

D. L. Hildenbrand and D. D. Cubicciotti, eds., *High Temperature Metal Halide Chemistry*, 1977.

R. G. Frieser and C. J. Mogab, eds., *Plasma Processing*, 1980.

K. K. Yee, *Int. Met. Rev.* **1**, 19 (1978).

K. C. Mittal, ed., *Adhesion Measurements of Thin Films, Thick Films and Bulk Coatings*, ASTM STP, Philadelphia, Pa., 1978, p. 640.

J. I. Duffy, ed., *Electroless and Other Non-electrolytic Plating Techniques*, Noyes Data Corp., Park Ridge, N.J., 1980.

W. Kern and G. L. Schnable, *IEEE Trans.* **ED-26**, 647 (1979).

G. Wahl and R. Hoffman, *Rev. Int. Temp. Refract. Fr.* **17**, 7 (1980).

R. S. Holmes and R. G. Loasby, *Handbook of Thick Film Technology*, Electrochemical Publications, 1976.

ASTM Metals Handbook, Vol. 2, American Society for Metals, 1964.

K. L. Mittel, ed., *Surface Contamination: Genesis, Detection and Control*, Vols. 1 and 2, Plenum Press, Inc., New York, 1979.

Journals

Thin Solid Films, Elsevier Sequoia SA, Lausanne, Switzerland.

J. Vac. Sci. Technol., American Vacuum Society.

Proceedings of the International Conference on Chemical Vapor Deposition, Electrochemical Society.

J. Electrochem. Soc.

Therm. Spray J., ASM International.

M. L. Thorpe, *Adv. Mater. Proc.*, 50–61 (May).

J. W. Dini, *Mater. Monu. Proc.* **4**(3), 331–346 (1989).

High Temperature Oxidation Resistant Coatings, Materials Advisory Board, National Academy of Science, Washington, D.C., 1970.

J. Lambert and co-workers, *Refractory Metals and Alloys*, Vol. 2, ASM International, 1990, pp. 557–585.

C. M. Cotell and co-workers, *ASM Surface Engineering Handbook*, Vol. 5, ASM International, 1984, pp. 856–863.

Merle Thorpe
Thorpe Thermal Technologies

REFRACTORY FIBERS

Fibrous materials may be naturally occurring or synthetically manufactured by thermal or chemical processes (Fig. 1) (see FIBERS, SURVEY). Refractory fibers are generally used in industrial applications at temperatures between 1000°C and 2800°C. These fibers may be oxides or nonoxides, vitreous or polycrystalline, and may be produced as whiskers, continuous filaments, or loose wool products.

Fiber chemistry determines whether the material is an oxide or nonoxide and can also influence its vitreous or polycrystalline physical form. Refractory fibers generally have diameters ranging from submicrometer to 10 μm, and lengths, as manufactured, may range from millimeters to continuous filaments.

Oxide fibers are manufactured by thermal or chemical processes into a loose wool mat, which can then be fabricated into a flexible blanket; combined with binders and formed into boards, felts, and rigid shapes; or fabricated into ropes, textiles and papers. The excellent thermal properties of these products make them invaluable for high temperature industrial applications.

Nonoxide fibers, such as carbides, nitrides, and carbons, are produced by high temperature chemical processes that often result in fiber lengths shorter than those of oxide fibers. Mechanical properties such as high elastic modulus and tensile strength of these materials make them excellent as reinforcements for plastics, glass, metals, and ceramics. Because these products oxidize at high temperatures, they are primarily suited for use in vacuum or inert atmospheres, but may also be used for relatively short exposures in oxidizing atmospheres above 1000°C.

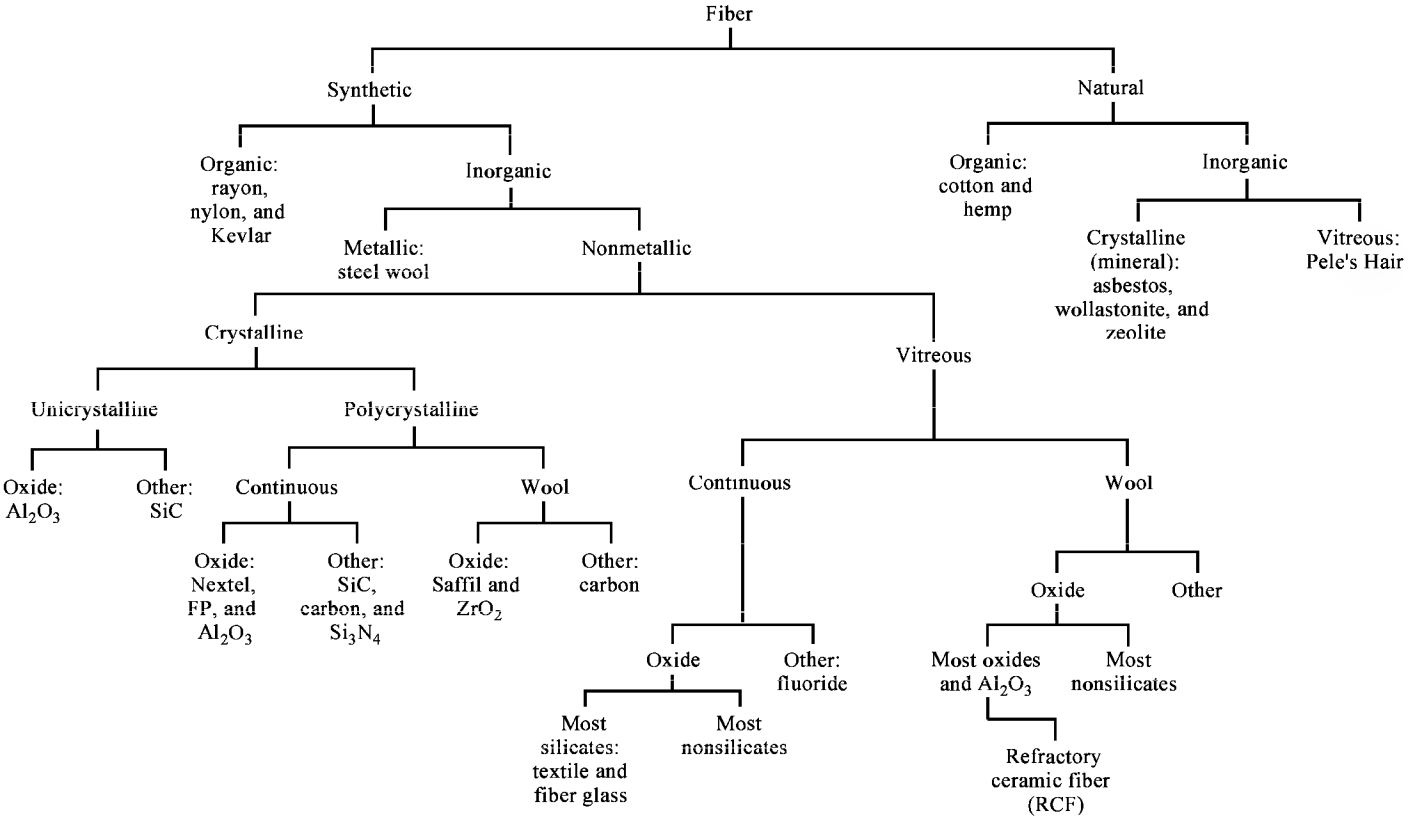


Fig. 1. Fiber classification tree.

Table 1. Maximum Use Temperatures of Refractory Fibers

Fiber type	Name	CAS Registry Number	Mp, °C	Use temperature, °C, max	
				a	b
Al ₂ O ₃	aluminum oxide	[1344-28-1]	2040	1540	1600
ZrO ₂	zirconium oxide	[1314-23-4]	2650	1650	1650
SiO ₂	silica	[7631-86-9]	1660	1060	1060
Al ₂ O ₃ -SiO ₂	aluminosilicate	[142844-00-6]	1760	1300	1300
Al ₂ O ₃ -SiO ₂ -Cr ₂ O ₃	chromia alumino-silicate	[142844-00-6]	1760	1430	1430
Al ₂ O ₃ -SiO ₂ -ZrO ₂	zirconia alumino-silicate	[142844-00-6]	1760	1430	1430
CaO-MgO-SiO ₂	calcium-magnesium silicate		1425	1000	1000
C	carbon	[7440-44-0]	3650	400	2500
B	boron	[7440-42-8]	1260	560	1200
BN	boron nitride	[10043-11-5]	2980	700	1650
SiC	silicon carbide	[409-21-2]	2690	1550	1800
Si ₃ N ₄	silicon nitride	[12033-89-5]	1900	1300	1800

^a Oxidizing atmosphere

^b Reducing atmosphere

Properties

Refractory fibers are most often used in applications above 1000°C. Table 1 shows the maximum long-term use temperatures in both oxidizing and nonoxidizing atmospheres. For short exposures, however, some of these fibers can be used with little degradation at temperatures within 100°C of their melting points.

The most important properties of refractory fibers are thermal conductivity, resistance to thermal and physical degradation at high temperatures, tensile strength, and elastic modulus. Thermal conductivity is affected by the material's bulk density, its fiber diameter, the amount of unfiberized material in the product, and the mean temperature of the insulation. Products fabricated from fine fibers with few unfiberized additions have the lowest thermal conductivities at high temperatures. A plot of thermal conductivity versus mean temperature for three oxide fibers having equal bulk densities is shown in Figure 2.

The effect of bulk density on thermal conductivity for silica and aluminosilicate fibers is shown in Figure 3. At relatively low densities, solid conduction of heat is negligible when considering the total heat transfer. This is because radiation is responsible for the majority of the heat passing through the insulation. Convection and gas conduction are also factors that affect heat transfer, but these are minor influences at the higher use temperatures. For example, at 1260°C, over 80% of the heat transfer is the result of radiation. ASTM C201 and ASTM C177 are the standard test procedures for measuring the thermal conductivity of refractory fiber materials.

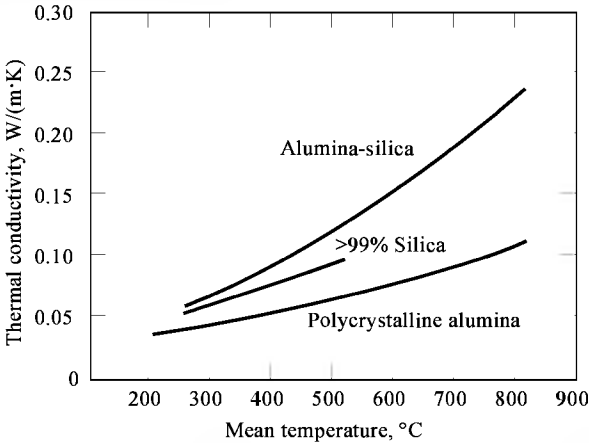


Fig. 2. Thermal conductivity of refractory fiber insulations with 96-mg cm³ density.

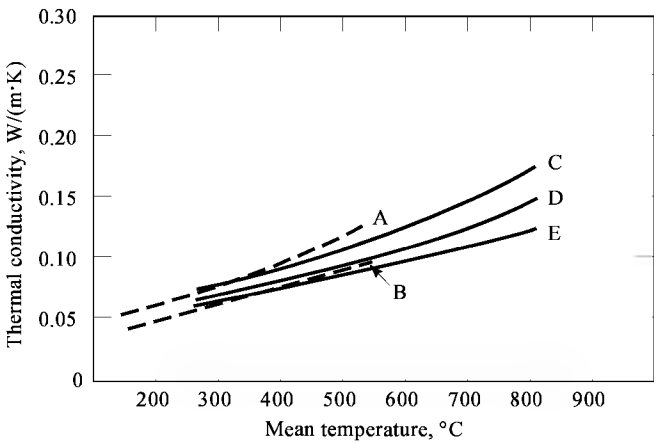


Fig. 3. Effect of density on thermal conductivity. A, 48-mg cm³ silica fiber; B, 96-mg cm³ silica fiber; C, 128-mg cm³ alumina–silica fiber; D, 192-mg cm³ alumina–silica fiber; E, 384-mg cm³ alumina–silica fiber.

For reinforcement, room temperature tensile strength and Young's modulus (stress–strain ratio) are both important. Typical values for refractory fibers are shown in Table 2.

Table 2. Mechanical Properties of Oxide and Nonoxide Fibers

Fiber type	Density, g cm ³	Tensile strength, GPa ^a	Young's modulus, GPa ^a
SiO ₂	2.19	5.9	72
Al ₂ O ₃	3.15	2.1	170
ZrO ₂	4.84	2.1	345
carbon	1.50	1.4	210
graphite ^b	1.66	1.8	700
BN	1.90	1.5	90
SiC ^b	3.21	2.0	480
Si ₃ N ₄ ^b	3.18	1.4	380

^a To convert GPa to psi, multiply by 145,000.

^b Single-crystal whiskers.

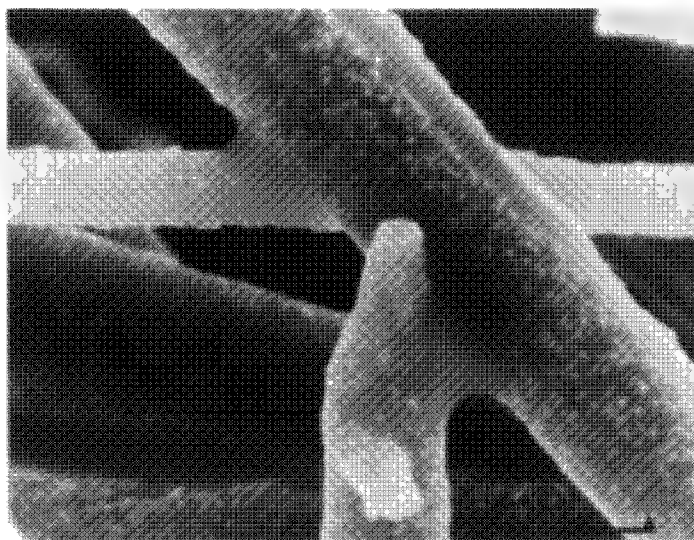


Fig. 4. Alumina-silica-chromia fiber after 120 h at 1426°C showing crystallization and sintering at contact points. Magnified $\times 5000$.

Heat treatment of vitreous refractory fibers often results in crystallization of the fiber (1). The rate of crystallization (qv), sometimes called devitrification, depends on temperature and time. Shrinkage and degradation of mechanical strength are attributable to this crystallization. Sintering, the bonding together of refractory fibers at temperatures below their softening points, is also responsible for shrinkage. At fiber contact points, solid-state diffusion of molecules joins the fibers together and transforms the previously flexible structure into a rigid mass. Figure 4 shows a scanning electron micrograph of alumina-silica-chromia fibers after exposure to a temperature of 1426°C for 120 h. The formation of mullite crystals in the vitreous fiber and the sintering at the fiber intersections are both visible. The diffusion of oxides in the fibers toward these contact points produces a shortening effect, which ultimately contributes to the overall shrinkage of the product. The property of linear shrinkage is an important design consideration for industrial furnaces and other high temperature applications.

Nonoxide Refractory Fibers

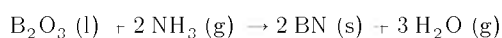
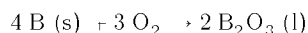
The most important nonoxide refractory fiber is silicon carbide. It was first produced as fibers or whiskers by a thermal chemical process in which one part silica was mixed with three parts carbon monoxide in an inert carrier gas at temperatures of 1300–1500°C. Silicon carbide fibers form in the cooler parts of the reactor tube (2). A less complicated technique has been developed by Corning Glass Works, in which silica and carbon monoxide gases are first formed by heating a mixture of carbon and silica (molar ratio of 2:1) in a hydrogen fluoride or hydrogen chloride atmosphere at temperatures of 1300 to 1550°C (3). The fibers grow more quickly through this process, which can be operated continuously. The resulting fibers consist of β -silicon carbide crystals surrounded by a silica sheath that prevents oxidation. A more recent innovation is the production of silicon carbide fibers by the pyrolysis of rice hulls in an inert or ammonia atmosphere (4).

Since 1960, a number of processes have been developed to produce high quality silicon nitride fibers. These fibers are a by-product in the production of silicon nitride powder from the reaction of silicon metal in a nitrogen atmosphere at high temperatures. Fiber yields can be increased by first reducing silica to silicon monoxide and then to silicon metal before the reaction with nitrogen (5). R represents a reducing agent in the following equations.



Any silicate that forms thermally and chemically stable residual compounds as its oxygen content is reduced provides a suitable source of silicon for this reaction. A typical process consists of alternating aluminum, silica, and graphite plates separated by 2–4-cm thick graphite spacers stacked in a graphite-lined alumina tube and heated to 1400°C for 12 h in a nitrogen atmosphere. After cooling for approximately 6 h the fibers are removed.

Boron nitride fibers are produced by nitriding boron filaments obtained during the chemical vapor deposition of boron on a heated tungsten wire (6,7). The tungsten wire is first passed through a reactor operating at 1100–1300°C and containing boron trichloride in a hydrogen carrier gas. The boron trichloride is reduced and boron is deposited onto the tungsten wire. To produce the boron nitride fiber, the boron filament is heated to 1000–1400°C in an ammonia atmosphere for approximately 6 h.



Nonoxide fibers in continuous form, eg, carbon, graphite, and boron, are often used in filament winding and in the manufacture of high strength, high modulus fabrics. Lightweight, high strength pressure vessels can be made by impregnating these fibers with epoxy resins and winding them onto plaster mandrels. After curing the resin, the mandrel is dissolved, leaving a thin-walled vessel. Carbon fibers are also used in sporting goods such as golf club shafts and tennis rackets as well as aerospace applications such as radar-transparent military aircraft, light and durable communications satellites, jumbo jets, and high temperature-tolerant rocket motor nozzles. Drill stems used in the collection of lunar samples are a combination of boron cloth with an epoxy-graphite filament. Carbon fibers can be used in nonoxidizing processes at temperatures up to 1800°C. Approximately 6.8×10^6 kg of carbon fiber is produced globally. Prices range from \$22 to more than \$220/kg for the specialty grades (8) (see CARBON AND GRAPHITE FIBERS).

Shorter, nonoxide filaments are primarily used as strength-enhancing reinforcements in resins, ceramics, and metals. These filaments have strengths of 1.4 to 20 GPa ($2\text{--}29 \times 10^5$ psi) and elastic moduli of 70 to 700 GPa ($1\text{--}10 \times 10^7$ psi). Silicon carbide whiskers (single-crystal fibers) are of special interest because they offer not only high strength and stiffness but also temperature resistance to 1800°C. Silicon carbide and silicon nitride fibers can be dispersed in a number of organic resins and then cast into shapes. These cast parts are used in high technology applications such as parts for specialty electronics, aircraft, and radomes. Boron nitride fibers are of particular value for the reinforcement of cast aluminum parts because the material can be wetted by molten aluminum. Silicon carbide and boron nitride are also used to reinforce gold and silver castings.

The growth and commercialization of the nonoxide fiber market parallels the high strength composite industry. If prices for nonoxide fibers with lengths of 2–10 cm reach the \$10–20/kg range, a large potential market should develop.

Oxide Fibers

High Purity Silica Fibers. In 1942, the Owens-Corning Fiberglass Corporation developed a process for leaching and refining E-glass fibers to give a >99% pure silica fiber (9). This fiber was widely used in jet engines toward the end of World War II. Shortly after the war, H. I. Thompson Co. began promoting silica insulating blankets for jet engines under the trade name Refrasil (10). In the 1950s, Micro-Quartz was developed by Glass Fibers, Inc. from a composition specifically designed for the leaching process (11).

In the 1990s, leached glass fibers having extremely fine diameters are produced using flame attenuation processing (see GLASS) (12). A glass composed of 75 wt% SiO₂ and 25 wt% Na₂O is melted in a typical glass furnace at 1100°C. Filaments having diameters of 0.3 mm are drawn from orifices in the bottom of the furnace. These filaments are passed through a gas flame and attenuated to a diameter of approximately 1.5 μm. The loose fibers are then subjected to an acid leaching process to remove the Na₂O, thoroughly rinsed, and dried. The resulting fibers have a purity of >99% silica.

Several techniques for manufacturing fused silica fibers by rod-drawing were developed in the early 1960s by Engelhard Industries and others (13). Fused silica rods having diameters of 6 to 7 mm are formed from molten, high purity silica and mounted in groups on a motor-driven carriage. Rod diameters are then reduced to 2 mm by forcing the rods through graphite guides and over a vertically oriented oxyhydrogen burner operating at 1800°C. The relatively large processed fibers then pass through a second graphite block where they are attenuated by an axial oxyhydrogen flame before being collected on a rotating drum. This process produces fibers that have diameters between 4 and 10 μm and silica contents of 99.95% (14). High thermal efficiency insulating felts having very low bulk densities produced by rod-drawing are essentially free of unfiberized material and can be used not only for jet engine insulation but also for space vehicles (see ABLATIVE MATERIALS). High purity silica fibers (Q Fiber) are used for the tiles of the reusable thermal protection system on the space shuttle.

In addition to aerospace uses, silica fibers can be twisted into sewing threads and yarns for weaving into fabrics. These fabrics are used extensively for heat-resistant clothing, flame curtains for furnace openings, thermocouple protection, and electrical insulation. The cloth can also be used to encapsulate other fibers to produce flexible sheets.

Chemically Produced Oxide Fibers. Refractory fibers from oxides of alumina or zirconia are difficult to manufacture by conventional melt technologies because of high melting points and low viscosities (see CERAMICS). In 1969, Union Carbide began to market the first 1650°C zirconia refractory fiber manufactured by a chemical precursor process (15). The precursor in this process is an organic fiber containing extremely small crystallites of polymer chains held together in a matrix of amorphous polymer. When immersed in a solvent, such as water, the fiber swells, expanding the spaces between the crystallites. Although a number of organic fibers, eg, wool, cotton, and cellulose acetate, have this swelling characteristic, rayon is preferred because of its structural uniformity and high purity. After an initial swelling and centrifugal dewatering, the rayon fiber is immersed in a 2-M aqueous solution of zirconyl chloride containing a small amount of yttrium salt. The excess solution is centrifuged, the fibers are dried, and the remaining fibrous material is fired at 400°C in an atmosphere containing <10 vol % oxygen. This treatment pyrolyzes the rayon to a carbonaceous residue. At the same time the zirconyl chloride is converted to microcrystalline zirconia fibers containing yttrium oxide for phase stabilization.

Another process for the manufacture of high temperature refractory fibers was developed by Imperial Chemical Industries (16). Both silica-stabilized alumina and calcia-stabilized zirconia have been manufactured by this sol-gel process. To produce an alumina fiber, a metallic salt such as aluminum oxychloride is mixed with a medium molecular weight polymer such as 2 wt % poly(vinyl alcohol). The solution is slowly evaporated in a rotary evaporator to achieve a viscosity of 80 Pa·s (800 P) and then extruded into fibers through a spinnerette. The fibers are collected and fired at 800°C to decompose the organic component. The resulting fine-grained alumina fibers have a porosity of 5 to 10 vol % and diameters of 3 to 5 μm. Fibers' inherent porosity makes them good for filtration and for use as catalytic substrates. For refractory applications the fibers are fired to 1400–1500°C for thermal stabilization and densification. Zirconia fibers can be produced using the same technology and with zirconium oxychloride, zirconium acetate, and calcia as the raw materials.

The 3M Company manufactures a continuous polycrystalline alumina-silica-boria fiber (Nextel) by a sol process (17). Aluminum acetate is dissolved in water and mixed with an aqueous dispersion of colloidal silica and dimethylformamide. This mixture is concentrated in a Rotavapor flask and centrifuged. The viscous mixture is then extruded through spinnerettes at 100 kPa (1 atm); the filaments are collected on a conveyor and heat-treated at 870°C to convert them to metallic oxides. Further heating at 1000°C produces the 10-μm diameter aluminum borosilicate fibers, which are suitable for fabrication into textiles for use at temperatures up to 1427°C.

Although more expensive than melt fiberization, the sol processes offer advantages in fiber chemistry selection. In melt fiberization, viscosity and surface tension are greatly influenced by additions of small quantities metallic oxides. In the sol process, where viscosity can be controlled independently, any number of metal salts may be added without adverse effects. These salts can serve as grain growth inhibitors, sintering aids, phase stabilizers, or catalysts.

Aluminosilicate Fibers. Vitreous aluminosilicate fibers, more commonly known as refractory ceramic fibers (RCF), belong to a class of materials known as synthetic vitreous fibers. Fiber glass and mineral wool are also classified as synthetic vitreous fibers, and together represent 98% of this product group. RCFs were discovered in 1942 (18) but were not used commercially until 1953. Typical chemical and physical properties of these materials are shown in Table 3.

RCF is produced by melting a combination of alumina and silica in approximately equal proportions or by melting kaolin clay in an electric resistance furnace (19,20). Trace ingredients, eg, zirconia, chromic oxide, and boria, may be added to enhance product properties (21). The molten mixture is formed into fiber either by blowing an air stream onto the molten material flowing from an orifice in the bottom of the furnace, or by directing the molten material onto a series of spinning wheels. The fibers are collected in a chamber as loose wool and are either packaged for future processing or converted into a blanket (mat) by a secondary needling heat-treating process. As produced, RCF typically contains 50–65 wt % fiber, with the remainder being unfiberized materials called shot (Fig. 5). Some shot is removed and recycled in the standard manufacturing process but additional separation steps may be required for certain applications.

Applications. RCF is primarily used in industrial applications where high temperature resistance, light weight, low thermal conductivity, and low heat storage are required. The light weight of RCF is particularly advantageous. Thermal properties are enhanced by the material's low bulk density. Lower capital costs can be realized because fewer structural furnace components are required. Compared to conventional hard refractories, installation times are reduced with these fibrous products. Low physical strength and chemical attack by ferrous metals and slags are disadvantages of RCF.

RCF is sold in a variety of forms, such as loose fiber, blanket, boards, modules, cloth, cements, putties, paper, coatings, felt, vacuum-formed shapes, rope, braid, tape, and textiles. The products are principally used for industrial applications as insulation in furnaces, heaters, kiln linings, furnace doors, metal launders, tank car insulation, and other uses up to 1400°C. RCF-consuming industries include ferrous and nonferrous metals, petrochemical, ceramic, glass, chemical, fertilizer, transportation, construction, and power generation incineration. Some newer uses include commercial fire protection and applications in aerospace, eg, heat shields; and automotive, eg, catalytic converters, metal reinforcement, heat shields, brake pads, and airbags.

Table 3. Typical Physical and Chemical Properties of Refractory Ceramic Fibers

Properties	Kaolin-base d RCF	High purity RCF	RCF with zirconia	RCF with chromia	Poly-crystalline alumina
		<i>Physical</i>			
color	white	white	white	blue green	white
maximum temperature rating, up to °C	1260	1316	1427	1427	1649
melting point, °C	1760	1760	1760	1760	1816
continuous use limit ^a , up to °C	1093	1177	1316	1371	1538
specific gravity ^b	2.56	2.65	2.65	2.65	3.3

specific heat, J (kg·K) ^c at 982°C	1089	1089	1089	1089	1047
fiber tensile strength, MPa ^d	1035				2000
fiber tensile modulus, GPa ^e	84				296
		Chemical ^f			
alumina	45	46	35	43	96
silica	53	54	50	54	4
zirconia			15		
chromium oxide				3	
other	1–2	trace	trace	trace	

^a This temperature is reported as a guide. Actual use limit depends on such factors as application, construction, fiber thermal stability, and anchoring system.

^b ASTM C135.

^c To convert J (kg·K) to Btu (lb·°F), divide by 4187.

^d To convert MPa to psi, multiply by 145.

^e To convert GPa to psi, multiply by 145,000.

^f Chemical analysis, nominal, % weight basis after firing.

Oxide and nonoxide refractory fibers have become essential materials for use in modern high temperature industrial processes and advanced commercial applications. Future process improvements, cost reductions, and performance enhancements are expected to expand the uses and markets for these specialized fibrous materials.

Health and Safety Factors. Concern has been expressed that synthetic vitreous fibers may pose a health hazard when inhaled. Research is ongoing, but epidemiological evidence, as of 1996, does not demonstrate a causal relationship between exposure to refractory ceramic fibers and the development of respiratory ailments, including cancer, fibrosis, and parenchymal disease (22). State-of-the-art inhalation toxicology research undertaken at the Research and Consulting Company (Switzerland) laboratories indicated that at extremely high doses (200 + times maximum recommended exposure levels) RCF is an animal carcinogen, but that a critical dose level may exist below which neither fibrosis nor tumors occur. The results of the toxicology research are under review. The International Agency for Research on Cancer (IARC) has classified all respirable synthetic vitreous fibers as 2B, a possible carcinogen.

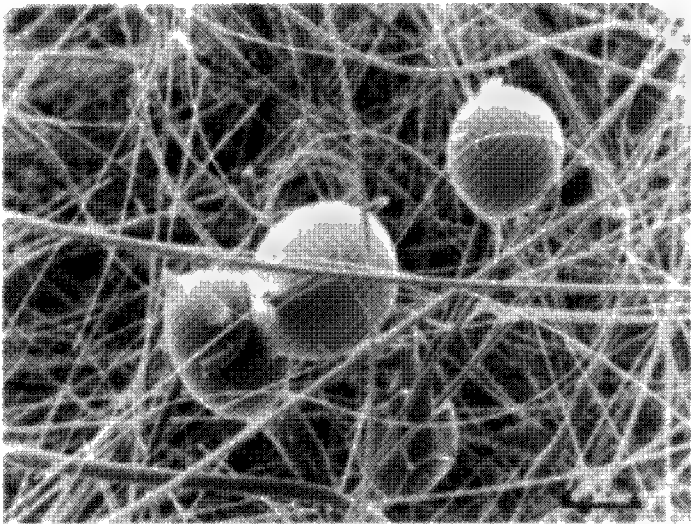


Fig. 5. Refractory fiber (1260°C) and unfiberized shot particles. Magnified × 1000.

Silica and aluminosilicate fibers that have been exposed to temperatures above 1100°C undergo partial conversion to mullite and cristobalite (1). Cristobalite is a form of crystalline silica that can cause silicosis, a form of pneumoconiosis. IARC has determined that cristobalite should be classified as 2A, a probable carcinogen. The amount of cristobalite formed, the size of the crystals, and the nature of the vitreous matrix in which they are embedded are time- and temperature-dependent. Under normal use conditions, refractory ceramic fibers are exposed to a temperature gradient, thus only the hottest surfaces of the material may contain appreciable cristobalite. Manufacturers' Material Safety Data Sheets (MSDS) should be consulted prior to handling RCF materials.

BIBLIOGRAPHY

"Refractory Fibers" in *ECT* 2nd ed., Vol. 17, pp. 285–295, by T. R. Gould, Johns-Manville Corp.; in *ECT* 3rd ed., Vol. 20, pp. 65–77, by W. C. Miller, Manville Corp.

1. L. Olds, W. Miller, and J. Pallo, *Am. Ceram. Soc. Bull.* **59**(7), 739 (1980).
2. U.S. Pat. **3,246,950** (Apr. 19, 1966), B. A. Gruber (to Corning Glass Works).
3. U.S. Pat. **3,371,995** (Mar. 5, 1968), W. W. Pultz (to Corning Glass Works).
4. U.S. Pat. **3,754,076** (Aug. 21, 1973), I. R. Cutler (to University of Utah).
5. U.S. Pat. **3,244,480** (Apr. 5, 1966), R. C. Johnson, J. K. Alley, and W. H. Warwick (to the United States of America).
6. U.S. Pat. **3,429,722** (July 12, 1965), J. Economy and R. V. Anderson (to Carborundum Co.).
7. U.S. Pat. **3,573,969** (Apr. 10, 1971), J. C. Millbrae and M. P. Gomez (to Lockheed Aircraft Corp.).
8. M. S. Reisch, *Chem. Eng. News*, **72**(48), 23 (Nov. 28, 1994).
9. U.S. Pat. **2,461,841** (Feb. 15, 1949), M. E. Nordberg (to Corning Glass Works).
10. *Chem. Eng. News* **25**(18), 1290 (1947).
11. U.S. Pat. **2,823,117** (Feb. 11, 1958), D. Labino (to Glass Fibers, Inc.).
12. U.S. Pat. **4,200,485** (Apr. 29, 1980), G. Price and W. Kielmeyer (to Johns-Manville Corp.).
13. U.S. Pat. **3,177,057** (Aug. 2, 1961), C. Potter and J. W. Lindenthal (to Engelhard Industries, Inc.).
14. Brit. Pat. **507,951** (June 23, 1939), (to W.C. Heraeus GmbH).

15.

U.S. Pat. 3,385,915 (May 28, 1968), B. Hamling (to Union Carbide Corp.).

16.

Brit. Pat. 1,360,197 (July 17, 1974), M. Morton, J. Birchall, and J. Cassidy (to Imperial Chemical Industries, Ltd.).

17.

U.S. Pat. 3,760,049 (Sept. 18, 1973), A. Borer and G. P. Krogseng (to Minnesota Mining and Manufacturing Co.).

18.

U.S. Pat. 2,467,889 (Apr. 19, 1949), I. Harter, C. L. Horton, Jr., and L. D. Christie, Jr. (to Babcock & Wilcox Co.).

19.

U.S. Pat. 2,714,622 (Aug. 2, 1955), J. C. McMullen (to Carborundum Co.).

20.

U.S. Pat. 3,066,504 (Dec. 4, 1962), F. J. Hartwig and F. H. Norton (to Babcock & Wilcox Co.).

21.

U.S. Pat. 3,449,137 (June 10, 1969), W. Ekdahl (to Johns-Manville Corp.).

22.

The Epidemiology and Toxicology of Exposure to Refractory Ceramic Fibers, RCF Coalition, Washington, D.C., Dec. 1993.

R. A. Waugh

Thermal Ceramics

REFRIGERATION

Refrigeration, as defined in this article, is the process of cooling materials using mechanical means and spans the temperature range of -157° to $+4^{\circ}\text{C}$. Human comfort cooling, which may be considered high temperature refrigeration in the temperature range of $4\text{--}13^{\circ}\text{C}$, is commonly referred to as air conditioning (qv). Ultra low temperature cooling from absolute zero, -273°C to -157°C , is commonly referred to as cryogenics (qv). This article focuses on refrigeration as defined above, specifically those applications using vapor-compression technology. Other systems capable of producing refrigeration, such as heat-activated machinery, ie, absorption systems, are not discussed.

Types of refrigeration may be categorized within five broad areas of application: (1) domestic appliances, ie, household refrigerator freezers; (2) commercial systems, ie, supermarket display cases, restaurants, and cafeterias; (3) cold storage and food processing, ie, refrigerated warehouses; (4) industrial, ie, liquefaction of gases, chemical process cooling, and crystallization; and (5) transportation, ie, refrigerated truck or trailer and marine containers. In 1991 it was estimated that annual investment in refrigeration machinery and equipment exceeded $\$100 \times 10^9$, and the value of the products treated by refrigeration exceeded $\$1 \times 10^{12}$ (1).

Basic Principles

All refrigeration systems are composed of five fundamental components: compressor, evaporator, condenser, expansion device, and refrigerant. A typical arrangement of these components is shown in Figure 1. The compressor raises the pressure of the refrigerant vapor so that its saturation temperature is above the temperature of the available cooling medium, ie, air or water. This difference in temperature allows transfer of heat from the vapor to the cooling medium so that the vapor can condense. The liquid flows through the expansion device, and in doing so a portion of the liquid flashes into a vapor, cooling the remaining liquid refrigerant below the temperature of the product to be cooled. This resulting difference in temperature allows heat to be transferred from the product to the refrigerant, causing the refrigerant to evaporate. The vapor formed must be removed by the compressor at a rate sufficient to maintain the low pressure in the evaporator and keep the refrigerant flowing. The resulting continuous flow process is referred to as the refrigeration cycle, shown schematically in Figure 2. This cycle is also called the reverse Rankine cycle and is governed by the principles of thermodynamics. For a review of these principles and the reverse Rankine cycle see the articles on [HEAT-EXCHANGE TECHNOLOGY](#) and [THERMODYNAMICS](#).

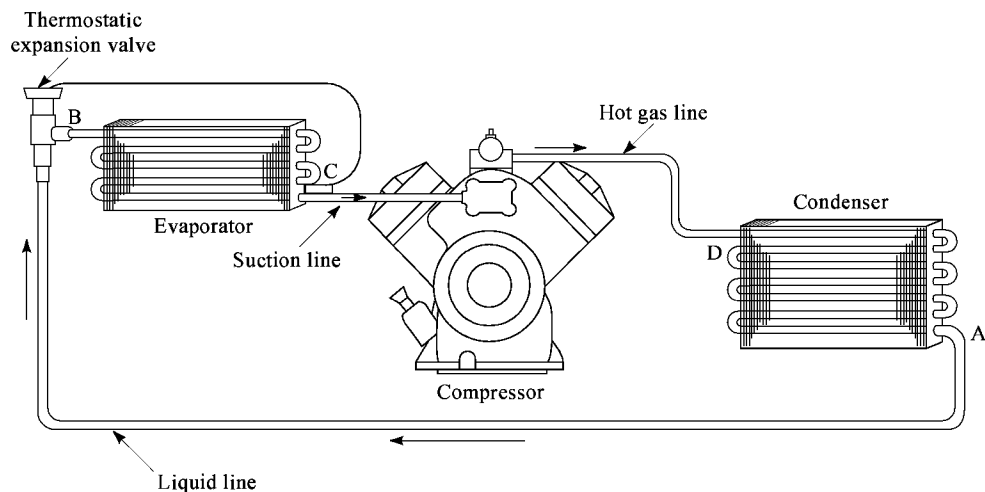


Fig. 1. Components of a refrigeration system: A, condenser outlet; B, evaporator inlet; C, evaporator outlet; D, compressor discharge.

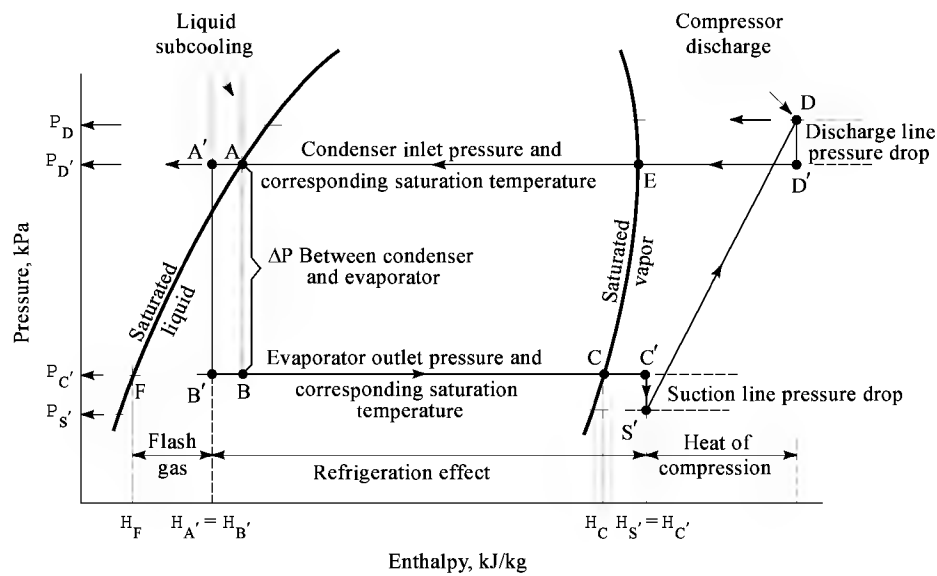


Fig. 2. Basic refrigeration cycle. A or A', condenser outlet; B or B', evaporator inlet; C or C', evaporator outlet; S', compressor inlet; D, compressor discharge; and D', condenser inlet. To convert kPa to atm, divide by 101.3. To convert kJ/kg to Btu/lb, multiply by 0.4302.

The standard unit of measure for refrigeration capacity is known as the refrigeration ton. It represents the amount of heat that must be removed from a short ton (909 kg) of water to form ice in 24 h. Its value is 3.51 kWt (12,000 Btu h(12.7 MJ h)). It is conventional to designate a kilowatt of refrigeration as a thermal kilowatt (kWt) to distinguish it from the amount of electricity (kWe) required to produce the refrigeration.

Refrigerant Nomenclature

The specific fluid used to produce refrigeration is a very important consideration. Historically, the refrigeration industry has used a wide variety of compounds that fall into three general categories: halocarbons, hydrocarbons, and inorganic fluids. Table 1 summarizes some of the key properties of these different types of refrigerants, including their normal boiling points, which establish the refrigeration temperature at which they could best be applied. Although not all of the possible refrigerant compounds are listed, Table 1 reflects the most commonly used ones in the five refrigeration categories. For saturated and unsaturated aliphatic hydrocarbons and aliphatic halogenated hydrocarbons designated as refrigerants, a numerical coding system has been developed that describes the refrigerant molecular structure by the use of a general designation having the form ABCD, in which A is the number of double bonds, B the number of carbon atoms less one, C the number of hydrogen atoms plus one, and D the number of fluorine atoms. The refrigerant dichlorodifluoromethane, CCl₂F₂, is designated R-12 (R refrigerant). Because there are no double bonds, A = 0; given one carbon, B = 1 – 1 = 0; given no H, C = 0 + 1 = 1; given two fluorines D = 2; and R-12 results from dropping the two initial zeros.

Table 1. Refrigerant Properties

Fluid number or name	Formula or mix, %	Mol wt	Normal bp, °C	T _c ^a , °C	P _c ^b , MPa	LFL ^c , %	Safety classification ^d
Fully halogenated chlorofluorocarbons							
CFC-13	CClF ₃	104.46	81.4	28.8	3.87		A1
CFC-115	CClF ₂ CF ₃	154.47	39.3	79.9	3.15		A1
CFC-12	CCl ₂ F ₂	120.91	29.8	111.8	4.11		A1
CFC-114	CClF ₂ CClF ₂	170.92	3.8	145.7	3.25		A1
CFC-11	CCl ₃ F	137.37	23.8	198.0	4.41		A1
CFC-113	CClF ₂ CCl ₂ F	187.38	47.6	214.4	3.46		A1
Hydrochlorofluorocarbons							
HCFC-22	CHClF ₂	86.47	40.8	96.2	4.99		A1
HCFC-124	CHClFCF ₃	136.48	12.1	122.5	3.63		A1
HCFC-142b	CClF ₂ CH ₃	100.50	9.3	137.1	4.25	6.9	A2
HCFC-123	CHCl ₂ CF ₃	152.93	27.9	183.8	3.67		B1
HCFC-141lb	CCl ₂ FCH ₃	116.95	32.0	204.7		7.4	A2
Hydrofluorocarbons							
HFC-23	CHF ₃	70.01	82.1	25.9	4.82		A1
HFC-32	CH ₂ F ₂	52.02	51.8	78.4	5.83	14.6	A2
HFC-125	CF ₃ CHF ₂	120.02	48.6	66.3	3.63		A1
HFC-143a	CF ₃ CH ₃	84.04	47.4	73.1	3.81	7.1	A2
HFC-134a	CF ₃ CH ₂ F	102.03	26.1	101.1	4.06		A1
HFC-152a	CHF ₂ CH ₃	66.05	24.2	113.3	4.52	3.7	A2
HFC-227ea	CF ₃ CHFCF ₃	170.03	17.3	101.9			NR
404A ^e (HFCs 125 143a 134a)	44 52 4	97.60	45.8	77.1	3.73		A1 A1
407C ^e (HFCs 32 125 134a)	23 25 52	86.20	43.6	86.1	4.62		A1 A1
410A ^e (HFCs 32 125)	50 50	72.50	50.5	72.5	4.96		A1 A1
507 ^f (HFCs 125 143a)	50 50	98.90	46.7	70.9	3.80		A3
Hydrocarbons							
propane (R-290)	CH ₃ CH ₂ CH ₃	44.10	42.1	96.8	4.25	2.1	A3
isobutane (R-600a)	CH(CH ₃) ₃	58.13	11.7	135	3.65	1.8	A3
n-butane (R-600)	CH ₃ (CH ₂) ₂ CH ₃	58.13	0.5	152.0	3.80	1.5	A3
Inorganic compounds							
ammonia (R-717)	NH ₃	17.03	33.3	133.0	11.42	15.0	B2
carbon dioxide (R-744)	CO ₂	44.01	g	31.1	7.37		NR

^a Critical temperature.

^b Critical pressure. To convert MPa to atm, divide by 0.1013.

^c Lower flammability limits (LFL) are expressed as vol % in dry ambient air. No entry means "None."

^d ASHRAE Safety Classification (2) (see Fig. 3). NR not rated.

^e The 400-series are zeotropic mixtures. T_c is based on bubble point.

^f The 500-series are azeotropic mixtures.

^g Sublimes.

Similarly, the saturated hydrocarbon propane, C₃H₈, is designated R-290, computed from A = 0, B = 3 – 1 = 2, C = 8 + 1 = 9, and D = 0.

If 90 is added to the designated R number, the resulting single digits individually reflect, in sequential order, the number of carbon, hydrogen, and fluorine atoms present, respectively. Chlorine is determined by difference, ie, the carbon valency of 4 minus number of H and F atoms. For example, R-12: 12 + 90 = 102, indicating 1 carbon, 0 hydrogen, 2 fluorine, and 2 chlorine atoms by difference.

Inorganic refrigerants are described somewhat differently. They are assigned three-digit numbers the first of which is 7, with the following two numbers comprising the molecular weight. Hence, NH₃ is designated 717 and SO₂ is designated 764.

Refrigerant mixtures are divided into two categories, azeotropes and zeotropes. The 500-series refrigerants are classified as azeotropes, since the vapor composition is identical to the liquid composition at a given pressure. The 400-series refrigerants are classified as zeotropes, because the equilibrium

vapor and liquid compositions are different and composition shifting can occur in a refrigeration cycle.

Refrigerants are placed into one of six safety categories, shown in Figure 3, per ASHRAE Standard 34-94 (2). Group A1 is classified as nonflammable and nontoxic; B1, as nonflammable, but slightly toxic; A2, as moderately flammable and nontoxic; B2, as moderately flammable and moderately toxic; A3, as highly flammable and nontoxic; and B3 as highly flammable and highly toxic. Refrigerants assigned to the various categories are shown in Table 1.

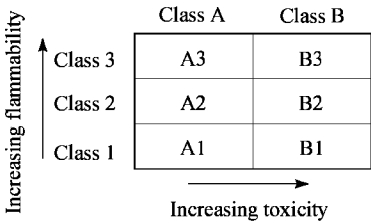


Fig. 3. Refrigerant safety group classifications established by ANSI ASHRAE 34-1992 (2). See text.

Environmental Factors

Within the family of halocarbon refrigerants are compounds that contain chlorine, fluorine, and carbon, the chlorofluorocarbons (CFCs). Hydrochlorofluorocarbons (HCFCs) contain chlorine, fluorine, carbon, and hydrogen. Halocarbons that contain only carbon, fluorine, and hydrogen are called hydrofluorocarbons (HFCs) (see FLUORINE COMPOUNDS, ORGANIC FLUORINATED ALIPHATIC COMPOUNDS). In the mid-1980s scientific investigations (3) confirmed that chlorine from compounds like CFCs and HCFCs were contributing to ozone depletion in the stratosphere, and a production phaseout schedule for these compounds was put in place in 1987 through the initialing of an international agreement known as the Montreal Protocol (4). Although originally scheduled for 50% production phaseout by the year 2000 in developed countries, the worsening ozone depletion has forced acceleration of the CFC phaseout. In the early 1990s, the CFCs were scheduled for production phaseout by January 1, 1996 and the HCFCs by 2020 or 2030, depending on the compound. Developing countries were generally given 10–15-yr extensions on the phaseouts. In the mid-1990s there was significant migration away from CFCs (and to a lesser extent, HCFCs) toward chlorine-free HFCs as fluids for refrigeration applications. Alternatives to CFCs and HCFCs must be sought and thus there has been increased interest in hydrocarbon and inorganic refrigerants such as propane and ammonia for applications where they have not traditionally been considered.

The selection of a refrigerant is based on many factors related to the specific application, including ozone depletion potential (ODP), global warming potential (GWP), toxicity, flammability, availability, cost, pressure, density, and theoretical cycle efficiency. Of the various factors, two are relatively new since 1985, ODP and GWP (Table 2). The term ODP refers to the relative ozone depletion potential of a refrigerant compared to that of CFC-11, which is arbitrarily assigned a value of 1.0. It is the numerical quantity describing the extent of ozone depletion calculated to arise from the release to the atmosphere of 1 kg of a compound relative to the ozone depletion calculated to arise from a similar release of CRC-11. The calculation is an integration of all known potential effects on the ozone layer over the entire time that traces of the compound would remain in the atmosphere. Therefore, HCFC-22 with an ODP of 0.05 has 5% the ozone depletion potential compared to that of CRC-11. This significantly reduced effect is one of the main reasons HCFCs are being phased out more slowly than CFCs. The term GWP is an index providing a simplified means of describing the relative ability of each greenhouse gas emission to affect radiative forcing and, thereby, the global climate. Radiative forcing reflects the factors that affect the balance between energy absorbed by the earth and the energy emitted by it in the form of long-wave infrared radiation. It is based on the gas infrared absorption characteristics and atmospheric lifetime. When written as GWP, the value is relative to carbon dioxide, CO₂, which is arbitrarily assigned a value of 1.0. CFC-11 has a GWP of 4000 on a molecule-to-molecule basis relative to CO₂. When written as HGWP, the value is relative to CFC-11, which is given a value of 1.0. HCFC-22 has an HGWP equal to 5% of CFC-11 on this basis.

Table 2. Refrigerant Toxicity and Environmental Data

Fluid number or name	TLV, ppm	ODP ^a	GWP ^b , 100 yr	Atmospheric life ^b , yr
Fully halogenated chlorofluorocarbons				
CFC-13	1000 ^c	na	11700	400
CFC-115	1000	0.52	9300	400
CFC-12	1000	1.0	8500	130
CFC-114	1000	0.8	9300	200
CFC-11	1000	1	4000	60
CFC-113	1000	1.07	5000	90
Hydrochlorofluorocarbons				
HCFC-22	1000	0.055	1500	15
HCFC-124	500 ^c	0.022	430	7
HCFC-142b	1000 ^c	0.065	1600	19
HCFC-123	10–30 ^c	0.02	93	2
HCFC-141lb	500 ^c	0.11	630	8
Hydrofluorocarbons				
HFC-23	1000 ^c	0	^d	310 ^e
HFC-32	1000 ^c	0	^d	6 ^e
HFC-125	1000 ^c	0	2500	28
HFC-143a	1000 ^c	0	2900	41
HFC-134a	1000 ^c	0	1200	16
HFC-152a	1000 ^c	0	140	2
HFC-227ea		0	3300	30
404A ^e (HFCs 125 143a 134a)	1000 ^c	0	3700	na
407C ^f (HFCs 32 125 134a)	1000 ^c	0	1300	na
410A ^f (HFCs 32 125)	1000 ^c	0	1900	na
507 ^g (HFCs 125 143a)	1000 ^c	0	3800	na
Hydrocarbons				

propane (R-290)	h	0	3 ⁱ	<1
isobutane (R-600a)	1000 ^c	0	3 ⁱ	<1
<i>n</i> -butane (R-600)	800	0	3 ⁱ	<1
<i>Inorganic compounds</i>				
ammonia (R-717)	25	0	<1	<1
carbon dioxide (R-744)	5000	0	1	120

^a ODP values are relative to R-11 GWP values relative to CO₂ and given for 100 years integrated time horizon. Atmospheric lifetimes are based on an *e*-folding decay.

^b Ref. 5.

^c Threshold limit value (TLV) not established by ACGIH. Value given is an estimate of a comparable index based on limited or incomplete toxicity testing made by chemical producers.

^d GWP is steady-state value.

^e Estimates based on limited data.

^f The 400-series are zeotropic mixtures.

^g The 500-series are azeotropic mixtures.

^h Listed as simple asphyxiant.

ⁱ GWP of the hydrocarbons results almost entirely from the GWP of the CO₂ owing to decomposition.

In the early 1990s a valuable concept in global warming assessment was developed called the Total Equivalent Warming Impact (TEWI) (6). Use of TEWI assesses the total warming impact on a system basis rather than on a molecule-to-molecule basis. In this concept, the equivalent CO₂ released to the atmosphere by leaking refrigerant (emissions) is combined with the CO₂ generated by fossil fuel combustion at the electrical power plant during production of electricity to drive the refrigeration system. In many cases the CO₂-equivalent from system emissions is only 5% of the total CO₂-equivalent from combined emissions and electrical power generation. Emissions are referred to as the direct global warming effect, and power plant CO₂ as the indirect effect. Thus, TEWI = direct effect + indirect effect, with units of kg-equivalent of CO₂. The concept of TEWI stresses the need to minimize emissions and maximize energy efficiency of systems and does not evaluate a refrigerant solely on its direct GWP or HGWP.

Household Refrigerator/Freezers

Domestic refrigeration in 1990 had a worldwide market of 56 × 10⁶ units annually (7). The vast majority of these used CFC-12 as their working fluid, and in general these systems would operate satisfactorily for 20 years or more. Over this period of time, an average compressor might operate 60–90 × 10³ h without failure.

There are several common layouts for refrigerator–freezers, eg, single-door refrigerators, side-by-side combinations, top-mount combinations, chest freezers, etc. Figure 4 gives a schematic of the refrigeration circuit. In such a system the typical design objective for performance might be to achieve average air temperatures of –19°C in the freezer section and –1°C in the general refrigerated space with continuous operation of the system in a 43°C ambient room. A system of ducts delivers proportional amounts of refrigerated air to the two compartments. The refrigerated air is derived from air blown through a fin-and-tube heat exchanger where the CFC-12 saturation temperature might be at –28°C and the entering air to the coil at –18°C.

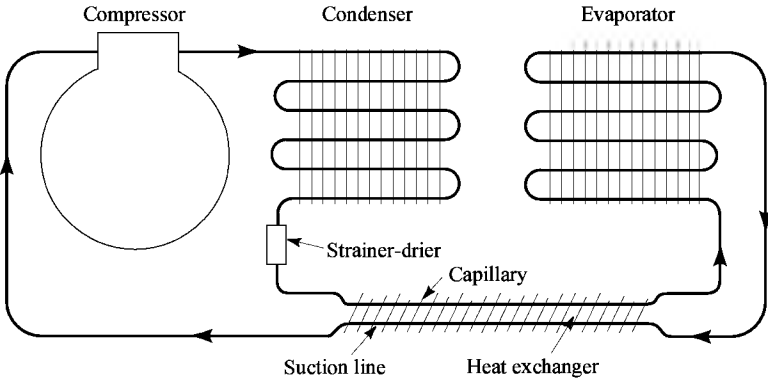


Fig. 4. Basic refrigerator circuit.

Compressors used in a typical refrigerator–freezer are of the positive displacement type, utilizing either reciprocating or rotary motion (Fig. 5) and range in capacities from about 90 kW (300 Btu h^{–1} 317 kJ h^{–1}) to about 600 kW (2000 Btu h^{–1}).

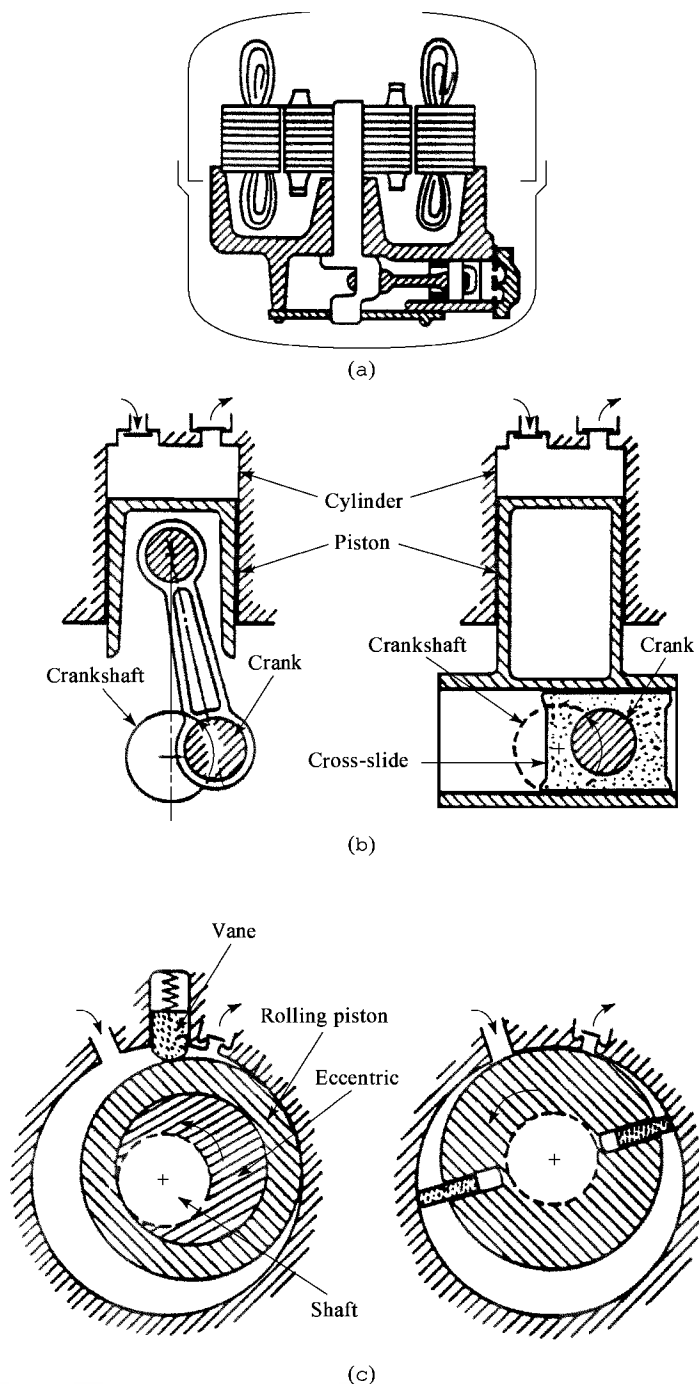


Fig. 5. Compressor types used in refrigerator freezers: (a) typical arrangement; (b) reciprocating piston mechanisms: connecting rod and Scotch yoke; and (c) rotary mechanisms: rolling piston and sliding vane.

Refrigerator freezers require considerable wall insulation to minimize heat loss, and CFCs have been used as the blowing agent for foam insulation used in these systems. The insulation is typically a rigid polyurethane foam blown with CFC-11 which has a very low vapor thermal conductivity of 0.0023 (k (J·m)⁻¹ (h·m²·K)⁻¹). The three basic components of polyurethane foam are an isocyanate, a polyol, and a blowing agent. The isocyanate and polyol are mixed, react exothermically, and the resulting heat boils the blowing agent. The resulting small bubbles froth the polyol-isocyanate mixture, which polymerizes, trapping the CFC-11 in the closed-cell structure (see FOAMED PLASTICS; URETHANE POLYMERS).

Whereas a typical refrigerator freezer might use 6 oz (170 g) of CFC-12 as refrigerant, it uses 2 lbs (0.91 kg) of CFC-11 in the blown foam. In the early 1990s, therefore, considerable research was initiated to find alternatives to these two CFCs for refrigerator freezers. The leading candidate to replace CFC-12 as a refrigerant is HFC-134a, which has similar, but not identical, operating pressures and thermodynamic properties. Whereas the CFCs are soluble in mineral oil used in the compressors, HFCs are not. Solubility of the refrigerant in the oil is necessary to ensure oil return to the compressor from oil entrained in the refrigerant during the compression process. The switch from CFCs to HFCs thus includes the shift from mineral oil to a synthetic oil known as a polyolester (POE), which is miscible with HFCs. Although HFC-134a is a leading candidate, it is not the only one, and significant research continues on a wide range of others (8). Likewise, considerable research is underway to identify effective alternatives to rigid foam blown with CFC-11 or devise other insulating concepts for refrigerator freezers (9).

Commercial Refrigeration

Nearly half of the retail food sold in the world consists of perishable or semiperishable foods that require refrigeration. The majority of commercial refrigeration systems are used in food merchandising and food service applications such as supermarkets, food stores, convenience stores, restaurants, cafeterias, and commercial and institutional kitchens. A consistent characteristic of commercial refrigeration applications is its use in buildings generally regarded as available for public use and having a potentially high occupancy rate. The system capacities typically range from 1 kW (0.25 t) to 500 kW (142 t). The design of commercial refrigeration systems and the type of refrigerant used depend on the desired operating temperatures. Fresh fruit and vegetables require air temperatures ranging from 0–13°C, fresh meat and dairy products require 2 to 2°C, and ice cream and frozen food 18 to 32°C.

The refrigerant CFC-12 has traditionally been used in medium (–15°C) to high (15°C) temperature refrigeration, while HCFC-22 has been used for temperatures down to –35°C, and R-502 down to –45°C. These three refrigerants represented 79, 2, and 19% of the refrigerants used for this application, respectively, in the years 1975–1995 (10).

As CFCs are phased out of production, it is anticipated that compounds such as R-22, R-134a, R-507, R-404a, which are more environmentally

benign, will replace them in commercial refrigeration applications.

Open self-service refrigerated display cases for medium and low temperatures are the most widely used type of refrigeration equipment in supermarkets. In recent years, glass door multideck models have gained popularity with store managers as competition between supermarkets has increased the need for more distinctive and attractive displays of refrigerated food. Table 3 shows the optimum storage temperature range for various products. Typically, food is to be sold within 24–48 h from the time displayed. Most retail refrigeration systems use reciprocating compressors in a single or multiplex configuration, with the system designed for at least a 10-yr life while operating at essentially 24 h d. In order to keep the refrigeration system operating within design guidelines it is important that the store temperature ambient conditions fall within certain limits. Refrigeration systems perform poorly when the store ambients exceed 24°C and 55° relative humidity because of evaporator coil frosting at low temperatures. This causes a significant loss in cooling capacity. To avoid this, modern stores are air-conditioned to maintain the appropriate range. Table 4 shows typical average store conditions over a period of one year.

Table 3. Temperatures in Display Refrigeration Systems

Fixture type	Temperature, ^a °C	
	Min	Max
meat, unwrapped display area	2	3.5
storage compartment	1	3
meat, wrapped display area	2	1
storage compartment	2	2
produce display area	2	7
storage compartment	2	7
dairy	2	6
frozen food	na	18
ice cream	na	24

^a Air temperature in refrigerated air stream.

Table 4. Typical Retail Food Store Conditions

Season	Ambient temperature		Moisture, g kg dry air	Relative humidity, °
	Dry bulb, °C	Wet bulb, °C		
winter	20.55	12.22	5.4	36
spring	21.11	14.44	7.9	50
summer	21.66	16.11	9.1	56
fall	21.11	14.44	7.9	50

Cold Storage and Food Processing

Refrigeration for large-scale cold storage and food processing involves very large refrigeration capacities (1000 kW) and a large refrigerant charge. Such systems are typically designed for a 20–30-yr service life. Historical usage of various refrigerants in cold storage warehousing is as follows:

Refrigerant	Total used, °
R-12, R-502	20
R-22	20
R-717 (ammonia)	60

The predominant use of ammonia in this application relates to several excellent thermodynamic properties. Liquid ammonia has a high specific heat, reasonable density and viscosity, plus high thermal conductivity. These factors make it a good heat-transfer fluid. Combining these factors with its high latent heat effect per unit weight and a high vapor density that provides a high volumetric capacity (kJ m³) makes it an efficient refrigerant, and less ammonia is required for a given system than other refrigerants. Therefore, system charge requirements are lower, which helps to reduce both initial and long-term operating costs. Since these types of systems can be installed in isolated areas with low human occupancy, the issue of toxicity and flammability associated with ammonia are manageable. Ammonia systems use heat exchangers constructed of noncopper materials, because copper and ammonia are not compatible. This tends to derate the ammonia systems somewhat since copper-based heat exchangers are more effective than noncopper heat exchangers.

Fluorocarbon-based refrigerants are nontoxic and nonflammable and require more charge, but they have lower volumetric capacities and require a larger system charge. However, they are also compatible with copper heat exchangers, which offsets the thermodynamic advantage of ammonia to some extent. In general, the initial and operating costs of fluorocarbon-based systems is slightly higher than ammonia equipment, but the additional investment may be justified in some cases to meet local codes or restrictions. For either type of refrigerant, the compression equipment is generally reciprocating or screw-type, with the screw being applied most commonly in systems of higher capacity (>30 kW).

As CFC refrigerants are phased out of production, their replacements in the cold storage applications will be HFC candidates such as R-134a, R-404A, and R-507. The use of ammonia, which is considered an environmentally benign refrigerant, will continue to play an ever-increasing role.

Industrial Refrigeration

Of all the application areas of refrigeration, the industrial sector uses the widest range of refrigerants and has the largest variety of cycles and systems, but uses the lowest total quantity of refrigerant compared to other areas of refrigeration.

Refrigeration can be accomplished in either closed-cycle or open-cycle configurations. In the closed-cycle, the refrigerant fluid is confined within the system and recirculates through the process in the cycle as shown in Figure 2. In the open-cycle, the fluid used as the refrigerant passes through the system once on its way to be used as a product or feedstock outside the refrigeration process, eg, the cooling of natural gas to separate and condense heavier components.

In addition to the distinction between open- and closed-cycle systems, refrigeration processes are also described as simple cycles, compound cycles,

or cascade cycles. Simple cycles employ one set of components and a single refrigeration cycle, as in Figure 2. Compound and cascade cycles employ multiple sets of components and two or more refrigeration cycles. The cycles interact to accomplish cooling at several temperatures or allow a greater span between the lowest and highest temperatures in the system than can be achieved with the simple cycle.

Closed-Cycle Operation. For a simple cycle the lowest evaporator temperature that is practical in a closed-cycle system is set by the pressure ratio capability of the compressor and by the properties of the refrigerant. Most high speed reciprocating compressors are limited to a pressure ratio of 9:1, so that the simple cycle is used for evaporator temperatures of $+2^{\circ}\text{C}$ to -50°C . Below these temperatures, the application limits of a single reciprocating compressor are reached. Beyond that, there is a risk of creating adverse conditions that cause excessive heat (which may break down lubricants), high bearing loads, excessive oil foaming at start-up, and inefficient operation because of reduced volumetric efficiency.

Centrifugal compressors with multiple stages can generate a pressure ratio up to 18:1, but their high discharge temperatures limit the efficiency of the simple cycle at these high pressure ratios. As a result, they have practical evaporator temperatures in the same range as reciprocating compressors (see FLUID MECHANICS; HIGH PRESSURE TECHNOLOGY).

By contrast, the single-screw and double-helical screw compressors (Fig. 6) can operate at high or low pressure ratios. They can be applied at pressure ratios below 2:1 and above 20:1 in a single stage. These compressors are suitable for use with higher pressure refrigerants such as CFC-12, HCFC-22, or ammonia.

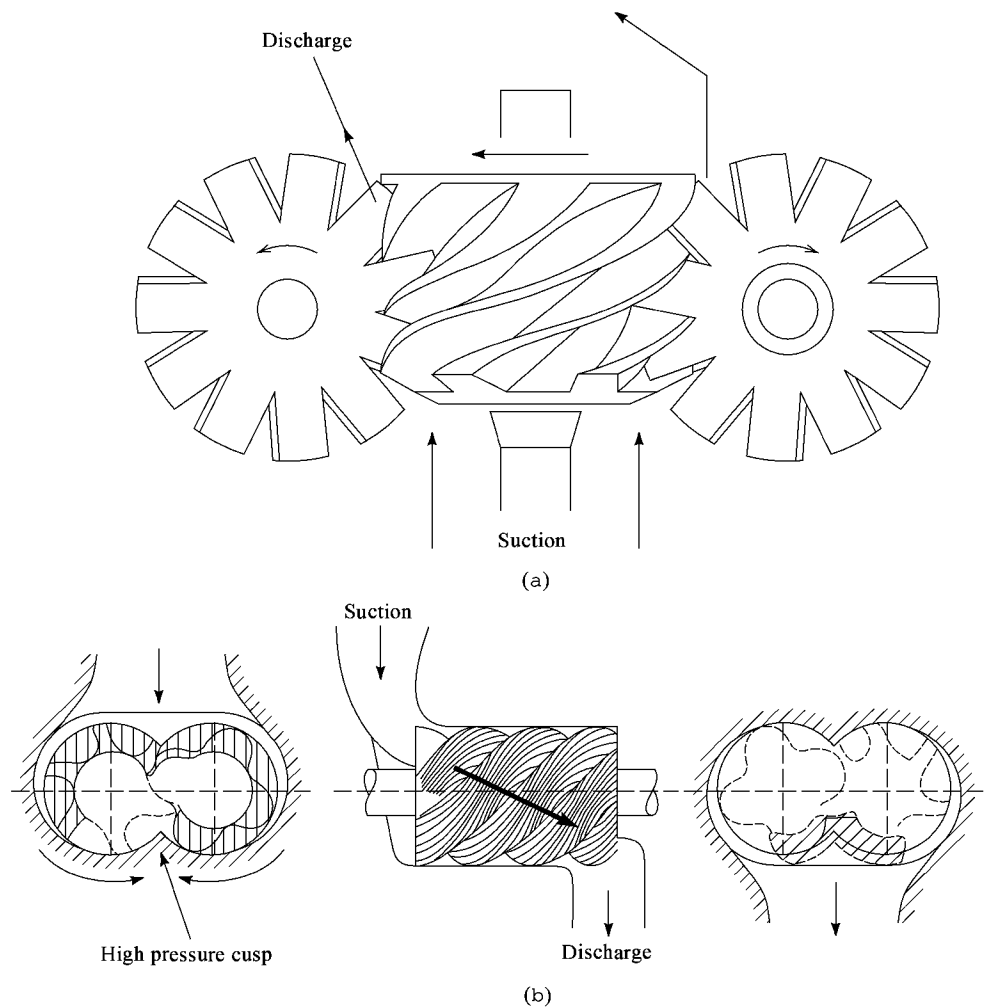


Fig. 6. (a) Single and (b) double-helical screw compressors (port areas shown as hatching).

The compound cycle (Fig. 7) achieves temperatures ca -100°C by using two or three compressors in series and a common refrigerant. This keeps the individual machines operating within their application limits. A refrigerant gas cooler is normally used between compressors to keep the final discharge temperature at a satisfactory level. Below -100°C , most refrigerants with suitable evaporator pressures have excessively high condensing pressures. For some refrigerants, the specific volume of refrigerant at low temperature may be so great as to require compressors and other equipment of uneconomical size. In other refrigerants, even though the specific volume of refrigerant is satisfactory at low temperature, the specific volume may become too small at the condensing condition.

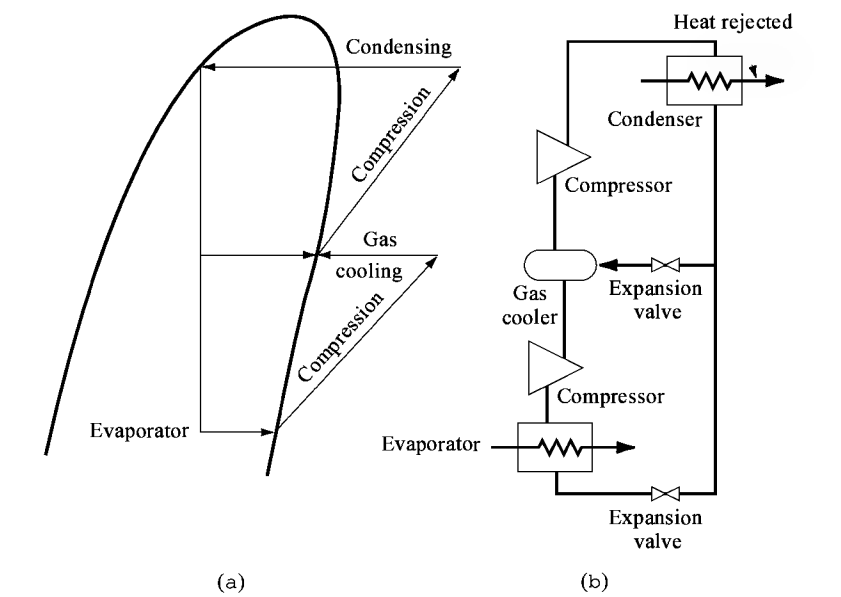


Fig. 7. (a) Pressure–enthalpy diagram and (b) system schematic for a compound cycle.

To satisfy these conditions, the cascade cycle is utilized (Fig. 8). This consists of two or more separate refrigeration cycles. The cascade condenser-evaporator rejects heat to the evaporator of the high temperature cycle, which condenses the refrigerant of the low temperature cycle. This makes possible the use of refrigerant such as R-13 (see Table 1) in the low stage, with pressure–temperature–volume (P–T–V) characteristics suited to low temperature. Refrigerants with P–T–V characteristics more favorable at higher temperatures, ie, R-12 or R-22, are used in one or more higher stages. For extremely low temperatures, more than two refrigerants may be cascaded as, for example, methane–ethylene–propane is to produce -160°C evaporator temperatures. Expansion tanks, sized to handle the low temperature refrigerant as a gas at ambient temperatures, are used during standby to hold pressures at levels suitable for economical equipment design.

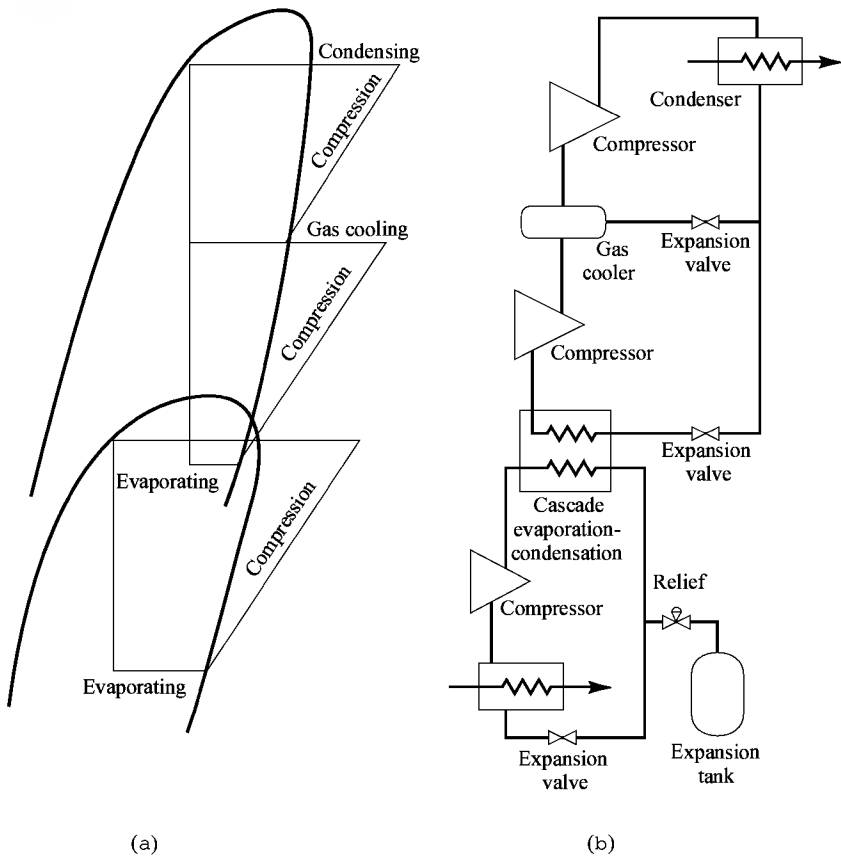


Fig. 8. (a) Pressure–enthalpy diagram and (b) system schematic for a cascade cycle.

In some cases, blade-type rotary compressors are used in low temperature applications as high volume, low stage, or booster compressor (Fig. 9). These booster compressors are applied at suction conditions varying from -87 to -20°C with compression ratios of 7:1 using CFC-12, HCFC-22, or ammonia.

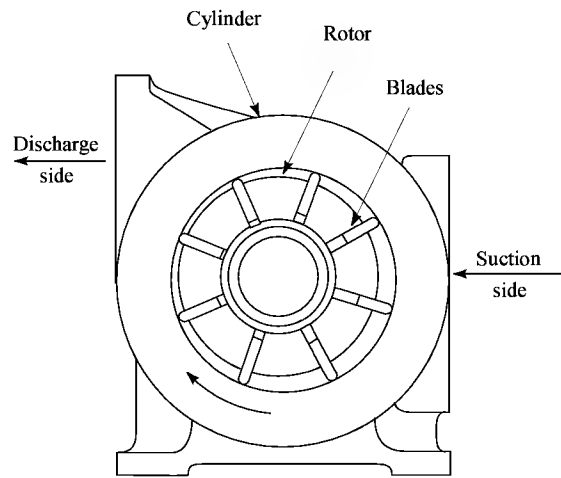


Fig. 9. Large rotary compressor.

Compound cycles using reciprocating compressors or any cycle using a multistage centrifugal compressor allow the use of economizers or intercoolers between compression stages. Economizers reduce the discharge gas temperature from the preceding stage by mixing relatively cool gas with the discharge gas before entering the subsequent stage. Either flash-type economizers, which cool refrigerant by reducing its pressure to the intermediate level, or surface-type economizers, which subcool the refrigerant at condensing pressure, may be used to provide the cooler gas for mixing. This keeps the final discharge temperature low enough to avoid overheating of the compressor and improves compression efficiency.

Compound compression with economizers also affords the opportunity to provide refrigeration at an intermediate temperature. This provides a further thermodynamic efficiency gain because some of the refrigeration is accomplished at a higher temperature, and less refrigerant must be handled by the lower temperature stages. This reduces the power consumption and the size of the lower stages of compression. Figure 10 shows a typical schematic with flash-type economizers. Process loads at several different temperature levels can be handled by taking the suction to an intermediate compression stage (Fig. 10a). The corresponding pressure–enthalpy (P–H) diagram (Fig. 10b) illustrates the thermodynamic cycle.

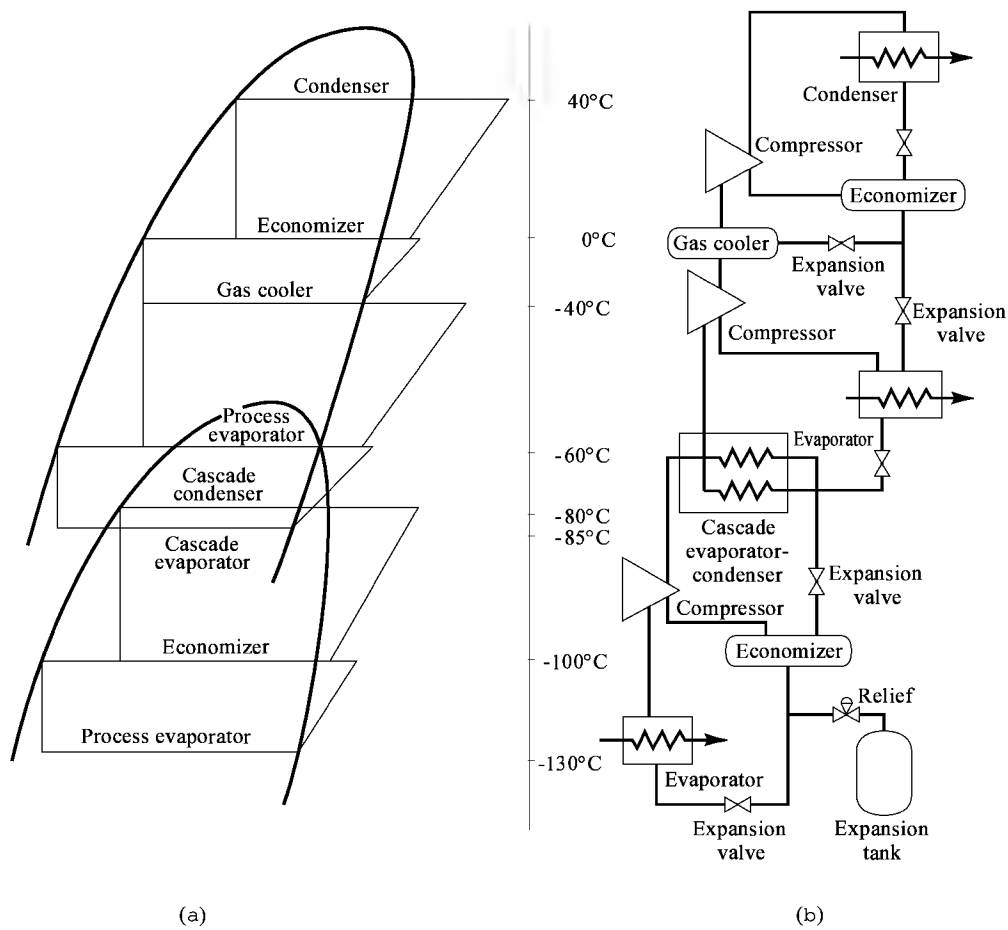


Fig. 10. (a) Pressure-enthalpy diagram and (b) system schematic for a system with flash economizers.

Flooded refrigeration systems are a version of the closed-cycle design that may reduce operating problems in some applications. In flooded systems, the refrigerant is circulated to heat exchangers or evaporators by a pump. Figure 11 shows the flooded cycle, which can employ any of the simple or compound closed-refrigeration cycles.

The refrigerant-recirculating pump pressurizes the refrigerant liquid and moves it to one or more evaporators or heat exchangers that may be remote from the receiver. The low pressure refrigerant may be used as a single-phase heat-transfer fluid as in A of Figure 11, which eliminates the extra heat-exchange step and increased temperature difference encountered in a conventional system that uses a secondary refrigerant or brine. This approach may simplify the design of process heat exchangers where large specific volumes of evaporating refrigerant vapor would be troublesome. Alternatively, the pumped refrigerant in the flooded system may be routed through conventional evaporators as in B and C, or special heat exchangers as in D. The flooded refrigeration system is helpful when special heat exchangers are necessary for process reasons, or where multiple or remote exchangers are required.

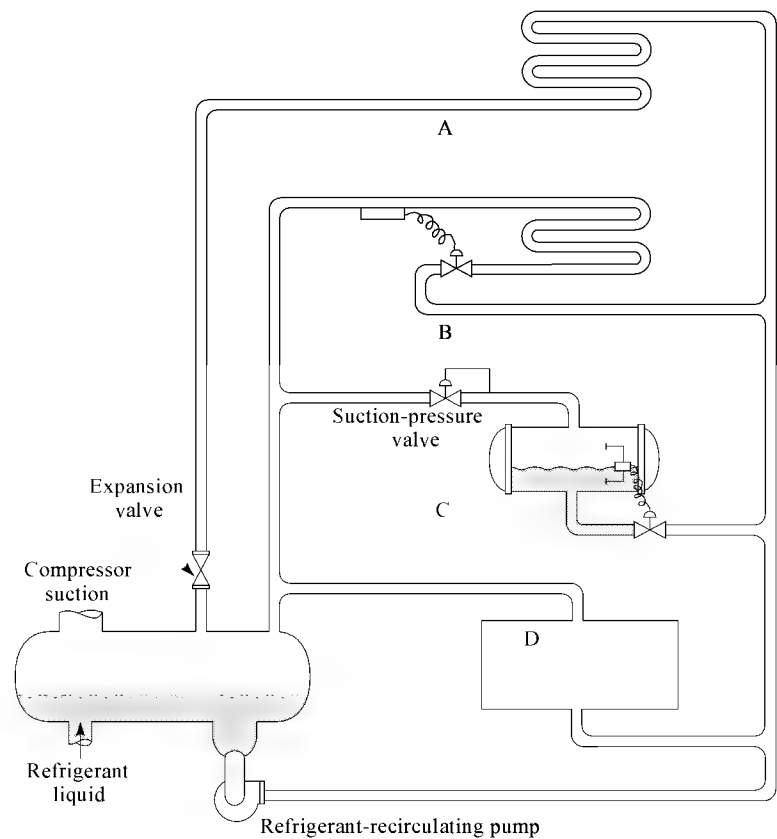


Fig. 11. Liquid recirculation in a flooded system where A refrigerant circulated through coils as brine, B direct expansion coils, C flooded evaporators, and D special heat-exchanger designs.

Open-Cycle Operation. In many chemical processes, the product to be cooled can itself be used as the refrigerating liquid. An example of this is in gathering plants for natural gas. Gas from the wells is cooled, usually after compression and after some of the heavier components are removed as liquid. This liquid may be expanded in a refrigeration cycle to further cool the compressed gas, which causes more of the heavier components to condense. Excess liquid not used for refrigeration is drawn off as product. A typical open-cycle is shown in Figure 12.

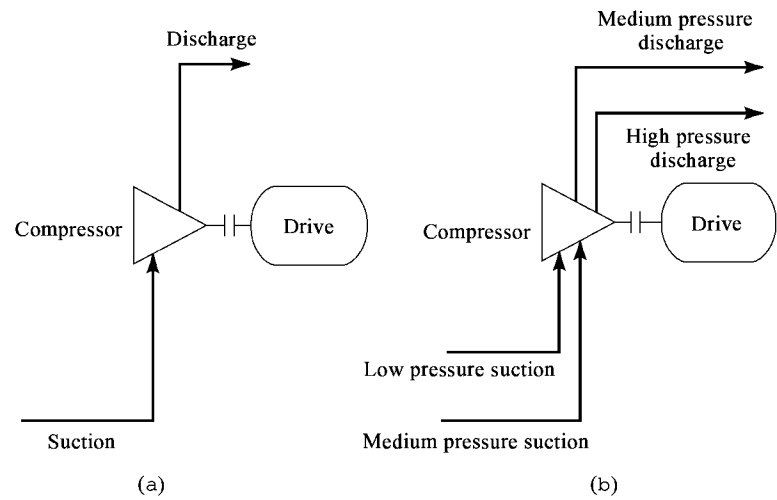


Fig. 12. Open-cycle compressors: (a) simple and (b) multiple product schematic.

Indirect Refrigeration (Brine). The process fluid is cooled by an intermediate liquid, water or brine, that is itself cooled by evaporating the refrigerant as shown in Figure 13. In the chemical industry, process heat exchangers must frequently be designed for corrosive products, high pressures, or high viscosities, and are not well suited for refrigerant evaporators. Other problems preventing direct use of refrigerant are remote location, lack of sufficient pressure for the refrigerant liquid feed, difficulties with compressor oil return, or inability to provide traps in the suction line to hold liquid refrigerant. The brine is cooled in the refrigeration evaporator and then pumped to the process load. The brine system may include a tank, either open or closed but maintained at atmospheric pressure through a small vent pipe at the top, or may be a closed system pressurized by an inert, dry gas.

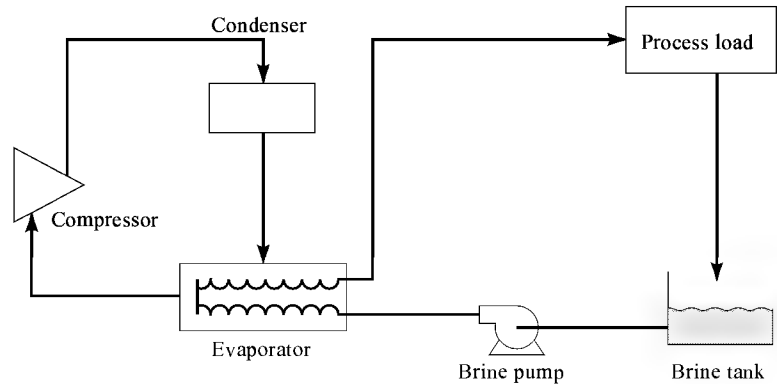


Fig. 13. Secondary brine system.

Working Fluids For Industrial Refrigeration. In 1991 the estimated refrigerant consumption for the industrial sector was 32 million kg distributed as follows: 16.6% CFCs, 43.3% HCFC-22, 33.3% NH₃, and 6.8% hydrocarbons.

In the future, CFCs and HCFCs will play a smaller role as they are phased out and replaced by HFCs such as R-134a, R-125, and R-507 where ammonia or hydrocarbons may not be considered appropriate. Because many industrial applications can safely manage the safety issues with ammonia, it is expected that the use of ammonia in this sector will continue to grow and may displace up to one-third of the industrial CFC and HCFC market of the late 1990s.

Transport Refrigeration

Refrigeration is used in trucks, truck trailers, intermodal containers, railcars, and ships. The majority of refrigerated vehicles fall into one of three refrigeration classifications: (1) 0 to 4°C, perishable produce; (2) -2 to 0°C, fresh meats; and (3) -43 to -17°C, frozen foods. Although the basic principles of the Rankine cycle apply to transport refrigeration, it has unique methods for powering the compression process.

The world fleet of truck and truck trailer refrigeration systems is estimated at 125,000 units, of which 30% are trailer units, 40% independent truck units, and 30% units driven off the truck engine. The container population is estimated at 300,000 units. Truck and truck trailer units have traditionally used CFC-12 or R-502, but in the 1990s some have switched to HCFC-22 or HFCs such as R-404a and R-507. Container units have typically used R-12, but some are supplied with HFC refrigerants such as R-134a.

Independent Engine- or Electric Motor-Driven. Many styles of independent engine- or electric motor-driven refrigeration units are available and are the most common for both truck and truck trailer. The one-piece, plug-type design is a self-contained unit that mounts in an opening provided in the front wall of the vehicle box. The condensing section is on the outside, and the evaporator on the inside. The two sections are separated by an insulating plug attached to the vehicle wall that supports the various parts of the refrigeration unit. For trailers, some of the plug units are tall and slender enough to mount between the tractor and trailer; others are short and deep and extend over the tractor roof. The auxiliary engines may operate on gasoline from the truck tank, from a separate gasoline or diesel fuel supply on trailers, or on liquefied petroleum gases. For intermodal containers, the most widely used refrigeration unit is electric motor-driven, with or without a companion engine-driven generator. The mounting plug for the unit essentially forms the entire end wall of the container. The intermodal container derives its name from the fact that with its independent refrigeration circuit it can be mounted in stacks on board ships for sea transport, off-loaded at a dock to a tractor trailer for road transport, or transferred by truck to a rail car for rail transport. They are designed to be stacked on top of one another for close storage of many units.

Self-contained, independent engine-driven plug-type units used in truck and truck trailers range in capacity from 6 to 12 kW to maintain trailer temperatures of 20°C in 38°C ambients. At trailer temperatures of -18°C in 38°C ambients, capacities range from 2–6 kW.

Power Take-Off From Engine or Transmission. This type of system is limited to trucks and there are several take-off means available. Most are some form of electric power generation equipment, belt-driven from the engine crankshaft, which produces either a regulated a-c voltage or rectified direct current for the compressor and fan motors in the body.

Lubrication System Considerations

Compressors move refrigerant around the system and require a lubricant to cool and reduce friction of moving parts. Although desirable, it is impractical in many instances to attempt complete isolation of the refrigerant; hence many refrigeration systems operate with a refrigerant–lubricant mixture in circulation. How much lubricant is in circulation depends on the system (11). Supermarket display cases, for example, may have lubricant circulations of 2–3% by weight. The need for solubility and miscibility stems from the recognition that once the lubricant gets out into the system it must be soluble in the refrigerant in order to get back to the compressor. As far as the compressor lubrication system is concerned, the most severe stress imposed on it results from the presence of the refrigerant. The refrigerant may be considered a contaminant in the sense that one of its main effects is the reduction of lubricant viscosity. Oil viscosity is an important design factor that is significantly affected by refrigerant dilution. For example, in the hydrodynamic lubrication of a simple coaxial shaft and sleeve bearing, the frictional power loss, *H*, owing to the shaft's rotating in the fluid is given by the following equation where *D* = shaft diameter, *C* = diameter clearance,

$$H = 2\pi^3 \frac{D^2}{C} Z \frac{N^2}{As}$$

Z = viscosity, *N* = RPM, and *As* = bearing area. Obviously the lower the operating viscosity, the lower the loss, but the design engineer must be concerned about the reduced viscosity of the oil after dilution by the refrigerant. Therefore, one of the first questions asked by design engineers is "What is the viscosity of the oil–refrigerant mixture as a function of temperature?" Once the viscosity–temperature–concentration relationship is known, the design engineer is able to determine the viscosity needs of the bearing system. If dilution by the refrigerant is expected, once a minimum viscosity for the bearing load is determined, a higher starting viscosity is usually selected to compensate for later dilution.

The lubricants used with CFC and HCFC refrigerants have traditionally been paraffinic, naphthenics, or aromatics. The paraffinics consist of all-straight or branched carbon-chain saturated hydrocarbons, such as isopentane. Naphthenics are saturated cyclic or ring structures such as cyclopentane. The aromatics are unsaturated cyclic hydrocarbons containing one or more rings characterized by alternate double bonds such as alkylbenzene. Although the CFCs and HCFCs are soluble and miscible with most of the paraffinic and naphthenic lubricants, some refrigerants such as HCFC-22 and R-502 show lower solubility with these oils at refrigeration temperature conditions. For improved solubility at low temperature, alkylbenzenes are frequently used with these refrigerants.

The newer HFC refrigerants are not soluble in or miscible with mineral oils or alkylbenzenes. The leading candidates for use with HFC refrigerants are polyol ester lubricants. These lubricants are derived from a reaction between an alcohol and a normal or branched carboxylic acid. The most common alcohols used are pentaerythritol, trimethylolpropane, neopentylglycol, and glycerol. The acids are usually selected to give the correct viscosity and fluidity at low temperatures.

Ammonia has low miscibility in mineral oils, alkylbenzenes, and polyol ester lubricants, particularly at low temperatures. A typical ammonia system uses a coalescing separator that removes all oil in droplet or aerosol form and drains it back to the compressor. Sometimes separators are equipped with some means of cooling the discharge gas to condense any oil that is discharged as a vapor.

BIBLIOGRAPHY

"Refrigeration" in *ECT* 2nd ed., Vol. 17, pp. 295–327, by J. R. Chambedain and R. A. Dorwart, York Division, Borg-Warner Corp.; in *ECT* 3rd ed., Vol. 20, pp. 78–107, by K. W. Cooper and K. E. Hickman, York Division, Borg-Warner Corp.

1. *Technical Options Report*, United Nations Environment Programme (UNEP) on Refrigeration, Air Conditioning, and Heat Pumps, Paris, France, 1991, p. 48.
2. *ASHRAE Handbook*, Fundamentals, American Society of Heating, Refrigerating, and Ventilating Engineers, Inc., Atlanta, Ga., 1993, p. 16.9.
3. M. J. Kurylo, *Proceedings of ASHRAE 1989 CFC Technology Conference at the National Institute of Standards and Technology, Gaithersburg, Md., Sept. 27–28, 1989*, ASHRAE Publications Department, Atlanta, Ga., 1990, pp. 5–15.
4. *Montreal Protocol, United Nations Environmental Program, Pts. I, II, III*, United Nations Ozone Secretariat, Nairobi, Kenya, 1994.
5. *1994 World Meteorological Organization (WMO)/NASA/UNEP Scientific Assessment Report No. 37*, University of Colorado Publications Services,

Boulder, Colo., pp. 1–28.

6. *Energy and Global Warming Impacts of CFC Alternative Technologies*, AFEAS Report, U.S. Department of Energy, Washington, D.C., 1991, pp. 3.9–3.13.

7. J. L. Boot, in Ref. 3, p. 75.

8. Ref. 7, p. 77.

9. I. R. Shankland, in Ref. 3, p. 87.

10. Ref. 1, Chapt. 4.

11. Y. Takaishi, *Seminar on New Technology of Alternative Refrigerants, Lubricants, and Materials Compatibility*, Japan Association of Refrigerants, Tokyo, Japan, 1993, pp. 101–106.

General Reference

ASHRAE Handbooks, 4 Vols. (Fundamentals (1993), Equipment (1992), Systems (1992), and Application (1995)), American Society of Heating, Refrigerating, and Ventilating Engineers, Inc., Atlanta, Ga., 1993.

Howard W. Sibley
Carrier Corporation



REGENERATED CELLULOSIC FIBERS

. See FIBERS, REGENERATED CELLULOSICS.

REGULATORY AGENCIES

Survey,
Chemical process industry,
Pharmaceuticals and cosmetics,
Power generation,

SURVEY

The purpose of Earth Day is the recognition that the environment, including all living things, must be protected. The first Earth Day, April 22, 1970, also marked the beginning of more concentrated attempts to develop regulations covering health, safety, and environmental issues; this effort is still underway in the 1990s.

Regulations change continuously with updates and reauthorizations, and the specifications of these regulations quickly become outdated. Therefore, the discussions in these articles on regulatory agencies provide only an introduction and summary of the U.S. Federal laws and regulations covering health, safety, and environmental issues. These articles should not be used in lieu of legal services. Current copies of the laws and regulations must be consulted in order to deal with specific situations.

The two main federal agencies involved in the protection of human health and the environment are the Environmental Protection Agency (EPA) and the Occupational Safety and Health Administration (OSHA). EPA's principal concern is the protection of the environment, in most cases, the area outside of an industrial facility. There are 10 regional offices that carry out the regulatory functions of the agency (Table 1). Primary laws covered by EPA are the Clean Air Act Amendments (CAAA), the Clean Water Act (CWA), Resource Conservation and Recovery Act (RCRA), Comprehensive Environmental Response, Compensation, and Liability Act (CERCLA), Toxic Substances Control Act (TSCA), and Federal Insecticide, Fungicide, and Rodenticide Act (FIFRA).

Table 1. Federal Regions for Various Regulatory Agencies, Including EPA and OSHA

Region	States, territories included	Location of regional office
1	Conn., Maine, Mass., N.H., R.I., and Vt.	Boston, Mass.
2	N.J., N.Y., P.R., and V.I.	New York, N.Y.
3	Del., D.C., Md., Pa., Va., and W. Va.	Philadelphia, Pa.
4	Ala., Fla., Ga., Ky., Miss., N.C., S.C., and Tenn.	Atlanta, Ga.
5	Ill., Ind., Mich., Minn., Ohio, and Wis.	Chicago, Ill.
6	Ark., La., N. Mex., Okla., and Tex.	Dallas, Tex.
7	Iowa, Kans., Miss., and Nebr.	Kansas City, Kans.
8	Colo., Mont., N.D., S.D., Utah, and Wyo.	Denver, Colo.
9	Ariz., Calif., Nev., Hawaii, Amer. Samoa, Guam, and N. Mariana Island	San Francisco, Calif.
10	Alaska, Idaho, Oreg., and Wash.	Seattle, Wash.

The principal function of OSHA is the protection of people, eg, employees, visitors, and temporary help, in the workplace. Regional offices are in the same areas as those of the EPA (see Table 1). The principal law covered by OSHA is the Occupational Safety and Health Act (OSHA).

There are a number of other federal agencies involved in related work. Pertinent agencies and their areas of concern are listed in Table 2. The control of the manufacture, use, and exposure to hazardous or toxic chemicals is mainly divided between EPA and OSHA. In addition, the Food and Drug Administration (FDA) has control over chemicals in food, drugs, and cosmetics. The Consumer Product Safety Commission (CPSC) is concerned with the safety of all consumer products, including child-resistant packaging regulations. The U.S. Department of Agriculture (USDA) maintains strict controls over chemicals in food. The specific concerns and requirements of these various laws are considered in detail in the regulatory agency articles pertaining to the specific industries. Additionally, the Department of Transportation (DOT) regulates the handling and transport of materials, including chemicals, by highway, rail, air, and water. Regulations cover training requirements for transporters, types of containers that can be used, labels and placards, incident reporting, etc (see TRANSPORTATION).

Table 2. Federal Agencies and Their Functions

Agency	Alias	Function
Agriculture Department	USDA	agriculture, animal and plant health inspection, forest service, food safety, Rural Electrification Administration, soil conservation service
Commerce Department	Commerce	Census Bureau, economics, trade, National Oceanic and Atmospheric Administration, Patent and Trademark Office
Department of Defense	DOD	Engineers Corps, dredge and fill operations, waterways, etc
Energy Department	DOE	all aspects of energy use, conservation, costs, etc
Health and Human Services Department	HHS	Food and Drug Administration, National Institute for Occupational Safety and Health (NIOSH)
Interior Department	Interior	land management, fish and wildlife, Geological Survey, mines, surface mining and reclamation
Justice Department	Justice	Antitrust Division, enforcement activities
Labor Department	Labor	employment standards and statistics, OSHA, Mine Safety and Health Administration (MSHA)
Transportation Department	DOT	Coast Guard; Federal Aviation Administration (FAA); highway, railroad, and maritime administration; hazardous material shipping
Treasury Department	Treasury	Alcohol, Tobacco, and Firearms Bureau (ATF), Customs Service, Internal Revenue Service
Council on Environmental Quality	CEQ	provides policy advice on environmental matters; implements National Environmental Policy Act (NEPA); coordinates environmental concerns among other agencies
Nuclear Regulatory Commission	NRC	ensures that radioactive materials are used and nuclear facilities are operated with regard for environment and public health, safety, and security
Office of Management and Budget	OMB	reviews regulations to ensure that they are cost-effective; approves paperwork or other information-gathering requirements

In addition to the federal agencies, there are many state agencies, as well as county and municipal agencies, that regulate the environmental and health areas. When state regulations are not as stringent as federal regulations, the federal rulings may take precedence. International laws and regulations must also be taken into account. International transportation requirements can affect shipments of products and raw materials. The use of products can also be affected by international rules.

The difference between laws and regulations is important. For the former, the U.S. Congress first passes a bill, which is then signed by the President, and thereby made a law or act. The act describes what Congress wants regulated, the general method to be used, and the ultimate results expected. It is then the responsibility of the designated agency to write and administer regulations to meet these requirements. First, the agency issues draft regulations that are mainly for internal review. It also tries to obtain the response of the affected portions of the public and industry. Next, proposed regulations are issued in *The Federal Register*, a daily document published by the General Services Administration. A specific comment period is allowed, during which public hearings may be held. The Office of Management and Budget also reviews the regulation, mainly to determine the financial impact of the regulations on industry. Once the comments are received and revisions are made, the final regulations are issued in *The Federal Register*. Every few years, Congress reviews a particular law and its regulations to see how well it is working. At that time, it can be amended, or reauthorized, to make any improvements.

The Federal Register contains, in addition to proposed and final regulations, notices for all of the federal agencies. Twice a year, a regulatory agenda is published in *The Federal Register*, listing all regulatory activities, from preproposed activities through proposed and final rulemaking. Expected dates for action are given.

Regulations are compiled and listed by agency in the *Code of Federal Regulations* (CFR). The listing first gives the number for the agency, then CFR, and then the part, or chapter, and number. An example is the EPA regulations on National Primary and Secondary Ambient Air Quality Standards: 40 CFR 50. Regulations of the EPA are listed under 40, with each set of regulations having a different chapter number. OSHA regulations are listed under 20.

The applicability of many of the government regulations are based on the Standard Industrial Classification (SIC) of the process and or facility. This is also known as the SIC Code and is assigned for tax and financial purposes. Because the classification can determine if a facility falls under certain regulations, it is important to ensure that the classification is correct. The SIC Code consists of four digits, the first two of which show the principal group. For chemicals and allied products, this group is 28. Table 3 lists the Industry Group Numbers under this major group.

Table 3. SIC Codes^a

Industry group number	Industry
281	industrial inorganic chemicals
2812	alkalies and chlorine
2813	industrial gases
2816	inorganic pigments
2819	industrial inorganic chemicals, not elsewhere classified
282	plastics materials and synthetic resins, synthetic rubber, cellulosic and other synthetic fibers, except glass
2821	plastics materials, synthetic resins, nonvulcanizable elastomers
2822	synthetic rubber, vulcanizable elastomers
2823	cellulosic synthetic fibers
2824	synthetic organic fibers, except cellulosic
283	drugs
2833	medicinal chemicals and botanical products
2834	pharmaceutical preparations
2835	<i>in vitro</i> and <i>in vivo</i> diagnostic substances
2836	biological products, except diagnostic substances
284	soap, detergents, and cleaning preparations; perfumes, cosmetics, and other toilet preparations
2841	soap and other detergents, except specialty cleaners
2842	specialty cleaning, polishing, and sanitation preparations
2843	surface-active agents, finishing agents, sulfonated oils, and assistants
2844	perfumes, cosmetics, and other toilet preparations
285	paints, varnishes, lacquers, enamels, and allied products
2851	paints, varnishes, lacquers, enamels, and allied products
286	industrial organic chemicals
2861	gum and wood chemicals
2865	cyclic organic crudes and intermediates, organic dyes, and pigments
2869	industrial organic chemicals, not elsewhere classified
287	agricultural chemicals
2873	nitrogenous fertilizers
2874	phosphatic fertilizers
2875	fertilizers, mixing only
2879	pesticides and agricultural chemicals, not elsewhere classified
289	miscellaneous chemical products
2891	adhesives and sealants
2892	explosives
2893	printing ink
2895	carbon black
2899	chemical and chemical preparations, not elsewhere classified

^a Ref. 1; under each industry classification, the manual lists specific chemicals that are included.

An important aspect of environmental, health, and safety laws and regulations is enforcement. Federal, state, and local regulatory authorities usually have large enforcement sections. In the environmental area, compliance audits are usually conducted annually. OSHA, both federal and state, usually audits based on a facility's accident incident rate.

Penalties for noncompliance are based on the severity of the violation, but is typically \$25,000 per day for each day of noncompliance. When a noncompliance is deemed to be willful, ie, the company knew they were committing a violation, the penalty can include a jail term in addition to the fine. When people or the environment suffer damage, a company can be sued as well.

BIBLIOGRAPHY

"Regulatory Agencies" in *ECT* 3rd ed., Vol. 20, pp. 108–127, by N. R. Passow, C-E Lummus.
1. *Standard Industrial Classification Manual*, Executive Office of the President, Office of Management and Budget, 1987; for sale by National Technical

Information Service, 5285 Port Royal Road, Springfield, Va., 22161, Order No. PB87-100012.

Nancy R. Passow
LONZA Inc.

CHEMICAL PROCESS INDUSTRY

The chemical process industry is highly regulated in the environmental, health, and safety area. Everything is affected, from the siting of a new facility to the transportation of raw materials and finished products, from the working conditions for employees to operating requirements for processes, packaging of finished goods, and dealings with the community. In addition to the regulatory requirements of government agencies, the chemical industry is developing standards of its own to ensure proper protection of the environment, employees, and the community. These include Responsible Care (registered by CMA); ISO14000: Environmental Management Systems; the sustainable development program of International Chamber of Commerce (ICC); and others.

The Chemical Manufacturers Association (CMA), the chemical industry's main trade association, developed the Responsible Care initiative. This program, which began in Canada, is a commitment on the part of the chemical industry to continuously improve health, safety, and environmental performance and to respond to public concerns. The initiative is based on 10 guiding principles and is implemented by six codes of management practices, covering Community Awareness and Emergency Response (CAER), Distribution, Employee Health and Safety, Pollution Prevention, Process Safety, and Product Stewardship. More than 35 countries around the world have adopted Responsible Care and are developing their own implementation programs.

In an attempt to bring uniformity to environmental protection, the International Standard Organization (ISO) followed up the ISO 9000 series of quality standards with the ISO 14000 environmental management standards. ISO 14001, Environmental Management Systems, is a public statement of environmental policy, which includes a commitment both to comply with relevant environmental legislation and a commitment to continual improvement; a planning process that identifies environmental objectives at all relevant levels within the company; designated management representatives to implement the company's plans; and procedures to identify and correct nonconformance, including periodic environmental management system audits. It is expected that, similar to ISO 9000, ISO 14001 certification will also be needed by companies wishing to do business in the international market.

Environmental Protection

Water. For a long time in the United States, the approach to water pollution control was through the establishment of water quality standards for receiving bodies of water, ie, rivers, streams, or lakes, with most limits established on a state-by-state basis. There was no effective, national, legal authority to limit the discharge of pollutants. In the late 1960s, the U.S. government revived an old law, the Rivers and Harbor Act of 1899 (the Refuse Act) (1). The law prohibited the discharge of anything into navigable waters unless a permit was obtained from the Corps of Engineers, thus providing a first step toward control of industrial discharges. This was followed by additional legislation, culminating in the passage of the Federal Water Pollution Control Act Amendments (FWPCA) of 1972 and the Clean Water Act (CWA) of 1977 (2). The objective of the FWPCA was to restore and maintain the chemical, physical, and biological integrity of the nation's waters.

Water Quality Standards. The first step in water quality standards is stream use classification. The individual states must decide what the uses of their water will be. The four categories, as defined by the EPA, are Class A, primary water contact recreation; Class B, propagation of desirable aquatic life; Class C, public water supplies prior to treatment; and Class D, agricultural and industrial uses. States may vary the definition of these classes to meet their own needs. The second step is to develop water-quality criteria. This is the specific concentration of a pollutant that is allowable for the designated use.

Effluent Guidelines and Standards. The Clean Water Act requires specific levels of control for dischargers. These are outlined in the Effluent Guidelines and Standards for various industrial categories. These standards limit the discharge of pollutants, usually in terms of a unit weight of pollutant per unit of either product or raw material, rather than a concentration in the discharge stream, in order to eliminate the use of dilution to meet limits.

The effluent standards require two levels of treatment: best practicable control technology (BPCT) and best available control technology (BACT). New standards covering additional industries are still being issued. Of special concern to the chemical industry are the following listed categories:

- Part 403: general pretreatment regulations for existing and new sources of pollution
- Part 411: cement manufacturing point source category
- Part 413: electroplating point source category
- Part 414: organic chemicals, plastics, and synthetic fibers (OCPSF)
- Part 415: inorganic chemicals manufacturing point source category
- Part 417: soap and detergent manufacturing point source category
- Part 418: fertilizer manufacturing point source category
- Part 419: petroleum refining point source category
- Part 422: phosphate manufacturing point source category
- Part 428: rubber manufacturing point source category
- Part 439: pharmaceutical manufacturing point source category
- Part 446: paint formulating point source category
- Part 447: ink formulating point source category
- Part 454: gum and wood chemicals manufacturing point source category
- Part 455: pesticide chemicals
- Part 457: explosives manufacturing point source category
- Part 458: carbon black manufacturing point source category

EPA has also developed pretreatment standards for industrial facilities that discharge directly to publicly owned treatment works (POTWs). The three types of pollutants of principal concern are pollutants that interfere with the operation of the POTW, pollutants that contaminate the sludges produced in the POTW, and pollutants that pass through the POTW or that are otherwise incompatible. One particular concern is volatile contaminants that can be stripped into the air during conventional wastewater treatment and become air pollution problems. These pretreatment standards are included in the effluent guidelines for the different industries.

National Pollutant Discharge Elimination System Permit Program. To ensure adherence to the effluent standards, the EPA developed the National Pollutant Discharge Elimination System (NPDES) permit program. Any source discharging or planning to discharge to any U.S. water must obtain an NPDES permit. The NPDES permit application must list all pollutants to be discharged, including any priority pollutants, and it must indicate the proposed treatment methods and resultant effluent. The NPDES permit indicates the allowable discharge and requires self-monitoring. As a minimum, the effluent standards must be met, but more stringent limits may be required. Where no standards for a pollutant or industry exist, limits are decided on an individual basis.

In recent years, storm water discharge has been added to the NPDES permit program. In addition to discharges generated within a facility, eg, from a process, storm water must be tested for contamination. Facilities must develop plans for minimizing storm water contact with chemicals on-site. For instance, outdoor chemical storage areas should be paved and diked to contain any storm water that could be contaminated as a result of spills. Loading and unloading areas must also be protected.

Other Laws. Many other laws have been passed to protect wetlands, coastal areas, and the oceans. The Safe Drinking Water Act (SDWA) (3) sets maximum contaminant levels (MCLs) for a variety of chemicals that can be present in drinking water. Although the greatest impact of this law is on providers of drinking water, there is an impact on the chemical industry. In order to protect underground sources of drinking water, the act contains a program regulating underground injection control (UIC) wells, a method used by industry for disposing of certain types of wastes.

Air. Studies have shown that 2500 years ago lead pollution caused by Greek and Roman silver smelters was a significant problem (4). Based on analysis of lake sediments and Greenland's ice, it was found that lead contamination from smelters in southern and central Europe was carried throughout the northern hemisphere. As long ago as the thirteenth century, air pollution has been linked to the burning of coal (4). The main concern was the smell from the sulfur in the coal and the effects of the soot. It was not until many years later that the effects of air pollution on people's health were discovered.

Various laws have been passed in the United States to control air pollution. The first law that had any real effect was the Clean Air Act of 1970 (CAA), which was followed by the Clean Air Act Amendments of 1977. Most recently, the Clean Air Act Amendments (CAAA) of 1990 (5) further changed and updated the requirements.

National Ambient Air Quality Standards. Under the Clean Air Act, six criterion pollutants, ie, pollutants of special concern, have been established by the EPA: sulfur oxides (SO_x), particulates, carbon monoxide (CO), nitrogen oxides (NO_x), ozone (photochemical oxidants), and lead. National Ambient Air Quality Standards (NAAQS) were developed by EPA based on threshold levels of air pollution below which no adverse effects could be experienced on human health or the environment.

The NAAQS are expressed in the form of ground level concentrations (GLC), which are the concentrations of pollutant in the ambient air as measured at ground level, in units of either micrograms per cubic meter or ppm. In order to convert a source's emission in kilograms per hour to a GLC, dispersion modeling must be used.

State Implementation Plans. The Clean Air Act requires a state implementation plan (SIP) in each state. The SIP indicates how that state intends to meet the NAAQS. State implementation plans can include such ideas as emission limitations, economic incentives or disincentives, plant closings or relocations, and changes in either operating schedules or methods. Although many states previously required sources to obtain permits, the CAAA of 1990 established a mandatory permit program. An issue of importance to the chemical industry is the complexity of amending a permit. Industry wants to have flexibility in revising permit conditions if plant operating conditions change.

Prevention of Significant Deterioration. EPA originally issued regulations for Prevention of Significant Deterioration (PSD) in December 1974 to protect clean air areas. Three air quality classes were designated: Class I to protect pristine areas, Class II to allow moderate development, and Class III to permit more intensive development. Most areas in the United States were initially designated as Class II. Many large national parks and wildlife areas have been classified as Class I.

As part of the PSD review, the applicant must show that BACT has been applied to all sources. Items to be evaluated include energy, environmental, economic, and other costs associated with each alternative technology as well as the associated benefits of reduced emissions. Another requirement is an ambient air quality analysis to show that the new emissions do not exceed either the NAAQS or PSD increments.

Nonattainment. EPA issued final rules for the Emission Offset Policy governing development in nonattainment areas. A new source must apply the lowest achievable emission rate (LAER) for the problem pollutant and must obtain a more than equivalent offsetting emission reduction from existing sources. Either the existing sources can be owned by the same company, or the reduction can be bought from other companies. In this way, new growth is allowed while air quality improvement is achieved.

Because nonattainment areas still exist, especially in urban areas, the 1990 CAAA contain new and more stringent requirements for such areas. The ambient air quality standards for ozone are of particular concern. Controls include tighter standards on emissions from motor vehicles, use of cleaner fuels, and additional controls on industrial facilities. One of the biggest impacts on the chemical industry is more stringent requirements for minimizing the emission of volatile organic compounds (VOCs). This can include process emissions as well as emissions from storage tanks.

Emission Standards. In order to have a nationwide basis for air pollution emission controls and to set a minimum emission limit, the EPA developed New Source Performance Standards (NSPS). The NSPS set specific pollutant emission limits or describe the best available control technology (BACT) that should be applied at that source. The EPA has issued NSPS, which apply to new construction as well as to large modifications, for many different sources. Sources in the chemical industry include the following.

- Sulfuric acid production units (Subpart Cb)
- Industrial–commercial–institutional steam-generating units (Subpart Db)
- Nitric acid plants (Subpart G)
- Sulfuric acid plants (Subpart H)
- Petroleum refineries (Subpart J)
- Volatile organic liquid storage vessels, including petroleum liquid storage vessels (Subparts K, Ka, and Kb)
- Phosphate fertilizer industry
 - wet process phosphoric acid plants (Subpart T)
 - superphosphoric acid plants (Subpart U)
 - diammonium phosphate plants (Subpart V)
 - triple superphosphate plants (Subpart W)
 - granular triple superphosphate storage facilities (Subpart X)
- Lime manufacturing plants (Subpart HH)
- Ammonium sulfate manufacture plants (Subpart PP)
- Asphalt processing and asphalt roofing manufacture (Subpart UU)
- Equipment leaks of VOC in the synthetic organic chemicals manufacturing industry (Subpart VV)
- Bulk gasoline terminals (Subpart XX)
- Rubber Tire Manufacturing Industry (Subpart BBB)
- Polymer manufacturing industry (Subpart DDD)
- Equipment leaks of VOC in petroleum refineries (Subpart GGG)
- Synthetic fiber production facilities (Subpart HHH)
- Synthetic organic chemical manufacturing industry air oxidation unit processes (Subpart III)
- Synthetic organic chemical manufacturing industry distillation operations (Subpart NNN)
- Petroleum refinery wastewater system VOC emissions (Subpart QQQ)

The 1990 CAAA expands the control of hazardous air pollutants (HAP). Previously, only a small number of hazardous air pollutants were regulated, under the National Emission Standards for Hazardous Air Pollutants (NESHAP). This has been expanded to cover a list of 189 pollutants, many of which are associated with chemical operations. Facilities are required to install maximum achievable control technology (MACT). MACT standards are issued by EPA for different industries and categories of sources. The applicability of these standards is based on a facility's potential to emit and whether or not it is a major source. Some categories of MACT standards in the chemical industry are Subpart F, for National Emission Standards for Organic Hazardous Air Pollutants from the Synthetic Organic Chemical Manufacturing Industry; Subpart G, for National Emission Standards for Organic Hazardous Air Pollutants from Synthetic Organic Chemical Manufacturing Industry Process Vents, Storage Vessels, Transfer Operations, and Wastewater; Subpart H, for National Emission Standards for Organic Hazardous Air Pollutants for Equipment Leaks; and Subpart I, for National Emission Standards for Organic Hazardous Air Pollutants for Certain Processes Subject to the Negotiated Regulation for Equipment Leaks.

Accidental Release Provisions. The 1990 CAAA includes provisions similar to OSHA's process safety management standard for minimizing the accidental release of air toxics. Based on types and quantities of hazardous chemicals on-site, a facility is required to develop and implement

risk management plans. The plans must be designed to prevent, detect, and respond to accidental hazardous chemical releases. In addition, facilities must provide this information to their local communities. The phrase that has been used to describe this information is worst-case scenarios.

Solid and Hazardous Waste. Regulation of pollution resulting from solid waste disposal was formulated at a much slower pace than regulation of air or water pollution. It was not until the Resource Conservation and Recovery Act (RCRA) of 1976 (6) was passed that substantial controls were authorized.

The main objectives of RCRA are to protect public health and the environment and to conserve natural resources. The act requires EPA to develop and administer the following programs: solid waste disposal practices providing acceptable protection levels for public health and the environment; transportation, storage, treatment, and disposal of hazardous wastes practices that eliminate or minimize hazards to human health and the environment; the use of resource conservation and recovery whenever technically and economically feasible; and federal, state, and local programs to achieve these objectives.

The section of the RCRA of most concern to the chemical industry is Subtitle C, the hazardous waste management regulations. The purpose of this section is to regulate hazardous wastes from their generation to their disposal. Facilities that generate, treat, store, or dispose of hazardous wastes are covered by these regulations.

The RCRA definition of solid waste covers a wide range of materials, including solid, liquid, semisolid, or contained gaseous material resulting from industrial, commercial, mining, and agricultural operations. A hazardous waste is a substance that must either be listed by the EPA or have a hazardous characteristic. Several types of solid wastes are specifically excluded from hazardous waste regulation because of their great volume or for other reasons. These include household wastes and some agricultural, mining, and fossil fuel combustion and exploration wastes.

A solid waste is considered hazardous if it is either a listed waste or a characteristic waste. Listed wastes include a list of specific processes that generate a waste and a list of discarded commercial chemical products. There are four hazardous waste characteristics: ignitability, corrosivity, reactivity, and toxicity. The last refers to the leachability of a waste and the resultant toxicity in the groundwater using the analytical method referred to as toxicity characteristic leaching procedure (TCLP). A list of substances included under TCLP is shown in Table 1.

Table 1. Maximum Concentration of Contaminants for Toxicity Characteristic

EPA hazardous waste number	Contaminant	CAS Registry Number	Regulatory level, mg/L
D004	arsenic	[7440-38-2]	5.0
D005	barium	[7440-39-3]	100.0
D018	benzene	[71-43-2]	0.5
D006	cadmium	[7440-43-9]	1.0
D019	carbon tetrachloride	[56-23-5]	0.5
D020	chlordan	[57-74-9]	0.03
D021	chlorobenzene	[108-90-7]	100.0
D022	chloroform	[67-66-3]	6.0
D007	chromium	[7440-47-3]	5.0
D023	<i>o</i> -cresol	[95-48-7]	4200.0
D024	<i>m</i> -cresol	[108-39-4]	4200.0
D025	<i>p</i> -cresol	[106-44-5]	4200.0
D026	cresol	[1319-77-3]	4200.0
D016	2,4-dichlorophenoxy	[94-75-7]	10.0
D027	1,4-dichlorobenzene	[106-46-7]	7.5
D028	1,2-dichloroethane	[107-06-2]	0.5
D029	1,1-dichloroethylene	[75-35-4]	0.7
D030	2,4-dinitrotoluene	[121-14-2]	30.13
D012	endrin	[72-20-8]	0.02
D031	heptachlor (and its epoxide)	[76-44-8]	0.008
D032	hexachlorobenzene	[118-74-1]	30.13
D033	hexachlorobutadiene	[87-68-3]	0.5
D034	hexachloroethane	[67-72-1]	3.0
D008	lead	[7439-92-1]	5.0
D013	lindane	[58-89-9]	0.4
D009	mercury	[7439-97-6]	0.2
D014	methoxychlor	[72-43-5]	10.0
D035	methyl ethyl ketone	[78-93-3]	200.0
D036	nitrobenzene	[98-95-3]	2.0
D037	pentachlorophenol	[87-86-5]	100.0
D038	pyridine	[110-86-1]	35.0
D010	selenium	[7782-49-2]	1.0
D011	silver	[7440-22-4]	5.0
D039	tetrachloroethylene	[127-18-4]	0.7
D015	toxaphene	[8001-35-2]	0.5
D040	trichloroethylene	[79-01-6]	0.5
D041	2,4,5-trichlorophenol	[95-95-4]	400.0
D042	2,4,6-trichlorophenol	[88-06-2]	2.0
D017	2,4,5-trichlorophenoxy (silvex)	[93-72-1]	1.0
D043	vinyl chloride	[75-01-4]	0.2

It is the generator's responsibility to determine if a substance is hazardous. Generator requirements include recordkeeping, labeling, using proper containers, providing information to transporters, following the manifest system, and periodic reporting to the EPA. The manifest system is a set of papers that is passed from the generator to the transporters of the waste, to those responsible for final disposal, and back to the generator to signify that the waste has been disposed of properly.

Regulations for owners and operators of hazardous waste treatment and storage and disposal facilities (TSDFs) cover both interim and general status. Any facility operating prior to November 30, 1980, is considered existing and is granted interim status if notification has been sent to EPA. The requirements cover operating methods and location, design, and construction of TSDFs. These include tanks; surface impoundments; waste piles; land treatment; landfills; incinerators; thermal treatment; chemical, biological, and physical treatment; and underground injection.

Groundwater and air quality monitoring are required for all facilities that have the potential to generate emissions. There are also requirements for contingency plans in the case of accidents, closure and post-closure plans, and financial requirements to ensure that closure plans can be followed. Permit applications must include an estimate of the composition, quantity, concentration, and frequency or rate of disposal, treatment, transport, or storage.

Regulations covering nonhazardous solid waste are mostly at the state and local level. Some states are starting to run out of space for landfills. As a

result, regulations are starting to increase for nonhazardous solid waste. Most of these are designed to encourage better management of solid waste, including source reduction, recycling, and reuse. One type of solid waste that is regulated separately is medical waste. Although this typically comes from hospitals and doctors' offices, some laboratory wastes from chemical plants, such as syringes used to inject samples into analytical equipment, may need to be handled specially.

RCRA also regulates underground storage tanks (USTs). USTs must be registered with a facility's designated state agency. Regulations include requirements for leak detection or inventory control system and tank testing; recordkeeping and reporting; corrective action; financial responsibility for corrective action and or third-party liability; and closure once the tanks are taken out of use.

One other law of interest is the Comprehensive Environmental Response, Compensation, and Liability Act (CERCLA) of 1980, known as the Superfund legislation (7). This act provides a means for the federal government to collect money from industry for use in cleaning existing and abandoned hazardous waste disposal sites and spills. 65% of industry's share of the fees comes from taxes on primary petrochemicals, 20% from taxes on organic raw materials, and 15% from taxes on crude oil. For operating facilities, there are regulations requiring reporting of spills and releases, based on reportable quantities (RQs), to the National Response Center (NRC). In addition, requirements for cleanup of facilities contaminated by past releases, as well as abandoned contaminated facilities, are included.

Emergency Planning and Community Right-to-Know Act. The Emergency Planning and Community Right-to-Know Act (EPCRA) (8), promulgated as part of the Superfund Reauthorization Act and originally known as SARA, Title III, requires facility emergency plans, spill and release reporting, annual inventory reporting, and annual toxic chemical emissions and pollution prevention reporting. The annual toxic emissions reports are called Toxic Release Inventory (TRI) reports, also known as Section 313 or Form R reports. These reports must include information on any on-site or off-site releases, disposal, treatment, or recycling of the listed toxic chemicals. Reporting requirements depend on the amount of the chemicals used or manufactured as well as the type of use. Sections 311 and 312 cover right-to-know requirements. The facility emergency plans and the annual inventory reporting must be shared with the community, through the local emergency planning committee (LEPC) and the state emergency response commission (SERC). The facility must also provide material safety data sheets (MSDSs) to the LEPC and other emergency responders.

Release Reporting. Both the Comprehensive Environmental Response, Compensation, and Liability Act (CERCLA) and EPCRA have requirements for reporting releases to the air, ground, or water. Lists of reportable chemicals or family of chemicals and their reportable quantity (RQ) have been issued (9). A reportable quantity is the amount, in pounds or kilograms, below which a release does not have to be reported. CERCLA requires only the reporting of releases from the CERCLA list; however, EPCRA requires reporting releases of both EPCRA- and CERCLA-listed substances. Reportable releases under CERCLA must be reported to the National Response Center, at (800) 424-8802. Reporting under EPCRA requires notifying the facility's LEPC (or relevant local emergency response personnel if there is no LEPC) and the SERC of any state likely to be affected. If a facility is near the border of another state, that state may have to be notified as well. Notification is required to be immediate, which is usually defined as within 30 minutes of the release. State or local authorities may have additional or different reporting requirements. Failure to report release in a timely manner can result in severe penalties from the regulatory authorities.

Multimedia Permitting. EPA has started looking at chemical plants with a multimedia view, ie, not just an air problem or a water problem, but how all aspects of a plant's operations, emissions, releases, etc, fit together. The driving force behind this is pollution prevention, not just treating wastes that are produced, but eliminating the production of the wastes in the first place. Of concern is the possibility of toxic use reduction (TUR) regulations. These require reduction in use, or even elimination, of chemicals that are deemed toxic. The chemical industry believes that with sufficient knowledge and proper controls, chemicals can be handled safely. Their focus is on reuse and recycling of chemicals in order to minimize emissions and releases.

Product Safety

Toxic Substances Control Act. EPA regulates the manufacture, use, and exposure to hazardous or toxic chemicals under a number of laws. For the chemical industry, the law of prime concern is the Toxic Substance Control Act (TSCA) (10), which was passed by the U.S. Congress in 1976. The two main goals of TSCA are acquisition of sufficient information to identify and evaluate potential hazards from chemical substances, and regulation of the production, use, distribution, and disposal of these substances.

One important aspect of TSCA is the premarket or import notification program. Before a manufacturer produces or imports a new chemical substance, EPA must be given 90 days' notice, the premanufacture notification (PMN). The notice includes all testing done by the manufacturer to determine the health effects of the chemical. Information on a facility's pollution prevention plans with regard to the manufacturing operation must also be included. EPA then has 45 days from the end of the review period to determine if the production or distribution of the chemical should be restricted or prohibited because the chemical presents an unreasonable risk. The manufacturer has an additional 30 days to object to EPA's decision. EPA has established an inventory of chemicals manufactured or processed in the United States (11). Any substance not on the list by August 30, 1980 is considered a new chemical and must be described in the premarket notification. Exemptions are given for substances registered as pesticides or regulated by the FDA.

As part of TSCA, EPA can require the testing of any chemical if there is the possibility of an unreasonable risk to health or environment or if there is significant human or environmental exposure. If the substance poses an unreasonable risk, EPA can prohibit the manufacture, processing, or distribution of the substance; limit the amount of the substance that can be manufactured, processed, or distributed; prohibit a particular use for the substance; limit the concentration of the substance during manufacture, processing, or distribution; regulate disposal methods for the substance; and require manufacturers to maintain records of process and to conduct tests to assure compliance with EPA rules.

Another section of TSCA requires the manufacturer to notify EPA if there is any indication of substantial risk from any chemical. Failure to do so by the manufacturer within a specified time period may result in civil penalties or possibly criminal prosecution.

TSCA also addresses the problem of polychlorinated biphenyls (PCBs) and chlorinated fluorocarbons (CFCs). EPA has developed regulations on the cleanup, handling, and disposal of PCBs. The manufacture and use of CFCs has been banned for all but essential uses, in accordance with the Montreal Agreement, an international treaty on worldwide use of CFCs.

Federal Insecticide, Fungicide, and Rodenticide Act. The Federal Insecticide, Fungicide, and Rodenticide Act (FIFRA) covers the use and manufacture of pesticides, rodenticides, etc, and includes such things as biocides and preservatives. Prior to manufacture, these regulated products must be registered with EPA. More details on these regulatory requirements are given in the articles FUNGICIDES, PESTICIDES, INSECT CONTROL TECHNOLOGY, and INDUSTRIAL ANTIMICROBIAL AGENTS.

Transportation. The U.S. Department of Transportation regulates the transport, packaging, and handling of hazardous substances, including chemicals. In addition, there are international laws and regulations. Substances are characterized based on their type of hazard and assigned identification numbers. The type of packaging that can be used for different types of hazards is regulated. For instance, there are only certain types of containers that can be used for flammables or corrosives. When hazardous materials are transported, the vehicle, eg, truck, tank truck, and tank car, must be placarded to show the hazard. The bill of lading must describe the material, including the hazards. The material safety data sheet (MSDS) must be available. In addition, there are many training requirements for those employees handling, packaging, or transporting hazardous chemicals. Only certified transporters may transport hazardous chemicals.

Employee Safety

Occupational Safety and Health Act. OSHA has broad responsibilities for protecting the workplace. The Occupational Safety and Health Act is administered by the Occupational Safety and Health Administration under the U.S. Department of Labor (12). The act covers all health and safety aspects of a worker's environment. Subpart Z of the Act, Toxic and Hazardous Substances, lists allowable employee exposure to many different chemical substances (13). These are given as ambient air concentrations over a certain time period, which usually is an 8-h time-weighted average. Sometimes a ceiling concentration is given as well. Certain substances, eg, vinyl chloride, benzene, and formaldehyde, are discussed in terms of necessary controls and limits. The monitoring of employee exposure is called industrial hygiene (qv).

In the Hazard Communication Standard, OSHA requires that all employees are trained in the hazards of the materials they are working with. This standard also requires that MSDSs be available for all hazardous chemicals at the worksite, accompany all shipments, and be sent to all customers. An

MSDS summarizes all of the important health, safety, and environmental information about a substance. A version of this standard also applies to laboratories, including research and development facilities, and requires the development of a laboratory hygiene plan.

Included in the OSHA regulations are standards for safe work practices such as lock-out tag-out and confined space entry, personal protective equipment, storage of hazardous materials, welding process, forklift operation, and requirements for fire protection. Basically, all activities within a chemical facility are covered by OSHA standards.

In response to a number of major incidents at chemical facilities, OSHA has issued process safety management standards. For a selected list of hazardous substances, a facility is required to have a process safety management program in place. This includes training to ensure employees know how to deal with hazards, analyses of processes using these hazardous substances, and management of change to ensure that any changes to equipment, processes, types of chemicals used, etc, are analyzed. Unlike EPA's accidental release provisions which focus on protection of the community, the focus of this standard is on protecting employees on-site.

The National Institute of Occupational Safety and Health (NIOSH), under the Department of Health and Human Services, works with OSHA. It is NIOSH's responsibility to determine safe exposure limits for chemical substances and to recommend to OSHA that these limits be adopted as standards.

Conservation, Land Use, and Other Areas Subject to Regulation

A new facility should be located in an area that is economically suitable (see PLANT LOCATION). At the same time, the social, environmental, and aesthetic effects must be considered. These decisions are part of land-use policies. As of this writing (1996), there is no national land-use program. However, there are certain laws that cover portions of this decision-making process and some states have such programs.

National Environmental Policy Act. The principal goal of the National Environmental Policy Act (NEPA) (14) is to establish a national policy that ensures continued growth and technological advancement while maintaining the quality of the environment. Any Federal agency sponsoring action that significantly affects the environment, eg, granting a construction permit or introducing new legislation, must issue a detailed report describing the environmental impact of the proposed action and alternatives. This report is called an environmental impact statement (EIS).

An EIS is written and issued by the federal agency involved with a particular project. The agency, however, relies on the owners of the proposed facility to provide the information contained in the EIS. The section on alternatives is considered the most important part of the EIS. The proposed project and its alternatives are usually described in detail. The environmental impacts of the proposal and the alternatives are presented. Based on all of this information, the federal agency determines if the proposed project is environmentally acceptable.

Acronyms in common use in the regulatory arena include the following:

Air quality control region (AQCR)
Advanced wastewater treatment (AWT)
Best available control technology (BACT)
Best available technology (BAT)
Best conventional pollutant control technology (BCT)
Biochemical oxygen demand (BOD)
Best practicable control technology (BPCT)
Clean Air Act (CAA)
Clean Air Act Amendments (CAAA)
Community Awareness and Emergency Response, under Responsible Care (CAER)
Comprehensive Assessment Information Rule, under TSCA (CAIR)
Confidential business information (CBI)
Comprehensive Environmental Response, Compensation, and Liability Act, also known as the Superfund Law (CERCLA)
Chlorinated fluorocarbon (CFC)
Code of Federal Regulations (CFR)
Chemical oxygen demand (COD)
Consumer Product Safety Commission (CPSC)
Clean Water Act (CWA)
Department of Transportation (DOT)
Environmental impact statement (EIS)
Environmental Protection Agency (EPA)
Environmental Planning and Community Right-to-Know Act (EPCRA)
Federal Insecticide, Fungicide, and Rodenticide Act (FIFRA)
Freedom of Information Act (FOIA)
Federal Water Pollution Control Act (FWPCA)
Good engineering practice (GEP)
Ground-level concentration (GLC)
Good laboratory practice (GLP)
Government Printing Office (GPO)
Hazardous air pollutant (HAP)
Lowest achievable emission rate (LAER)
Local Emergency Planning Committee (LEPC)
Leaking underground storage tanks (LUST)
Maximum achievable control technology (MACT)
Material Safety Data Sheet (MSDS)
National Ambient Air Quality Standard (NAAQS)
National Environmental Policy Act (NEPA)
National Emission Standards for Hazardous Air Pollutants (NESHAP)
National Institute for Occupational Safety and Health (NIOSH)
Nitrogen oxides (NO_x)
National Pollutant Discharge Elimination System (NPDES)
Notice of Proposed Rulemaking (NPRM)
National Response Center or Nuclear Regulatory Commission (NRC)
New Source Performance Standard (NSPS)
Occupational Safety and Health Act (and Administration) (OSHA)
Polychlorinated biphenyl (PCB)
Permissible exposure limit (PEL)
Premanufacture notification, under TSCA (PMN)
Publicly owned treatment work (municipal wastewater treatment facility) (POTW)
Potentially responsible party, under Superfund (PRP)
Prevention of Significant Deterioration (PSD)

Resource Conservation and Recovery Act (RCRA)
Reportable quantity (RQ)
Superfund Amendments and Reauthorization Act (SARA)
Safe Drinking Water Act (SDWA)
State Emergency Response Commission (SERC)
Standard Industrial Classification Code (SIC)
State implementation plan (SIP)
Sulfur oxides (SO_x)

Significant new use rule, under TSCA (SNUR)
Synthetic organic chemical industry (SOCMI)
Spill prevention control and countermeasure plan (SPCC Plan)
Toxicity characteristic leaching procedure (TCLP)
Threshold planning quantity (TPQ)
Toxic Release Inventory (TRI)
Toxic Substances Control Act (TSCA)
Treatment, storage, and disposal facility (TSDF)
Total suspended particulates (air) (TSP)
Total suspended solids (wastewater) (TSS)
Time-weighted average (TWA)
Underground-injection controls (UIC)
Underground storage tank (UST)
Volatile organic compounds (VOC)

BIBLIOGRAPHY

"Regulatory Agencies" in *ECT* 3rd ed., Vol. 20, pp. 108–127, by N. R. Passow, C-E Lummus.

1. *The Rivers and Harbor Act of 1899*, 33 USC §§401–413, and Executive Order 11574.
2. *Clean Water Act of 1977*, 33 USC §§1251–1387; 33 CFR Parts 320–330, 335–338; 40 CFR Parts 104–140, 230–233, 401–471.
3. *Safe Drinking Water Act*, 42 USC §§300f–300j–26; 40 CFR Parts 141–149.
4. "Ancient Silver Smelters Polluted the Hemisphere", *The Record*, Sept. 23, 1994, p. A-23; and P. Brimblecombe, "Attitudes and Responses Towards Air Pollution in Medieval England," *Air Poll. Contr. Assoc.* (Oct. 1976).
5. *Clean Air Act*, 42 USC §7401, *et seq.*; 40 CFR Parts 50–87.
6. *Resource Conservation and Recovery Act* of 1976, 42 USC §§6901–6992; 40 CFR 240–281.
7. *Comprehensive Environmental Response, Compensation, and Liability Act*, 42 USC §§9601–9675; 40 CFR Parts 300–311.
8. *Emergency Planning and Community Right-to-Know Act*, 42 USC §§11001–11050; 40 CFR Parts 350, 355, 370, 372.
9. *Comprehensive Environmental Response, Compensation, and Liability Act*, (CERCLA) Reportable Releases: 40 CFR 302, "EPA Designation, Reportable Quantities, and Notification Requirements for Hazardous Substances under CERCLA," Table 302.4, "Lists of Hazardous Substances and Reportable Quantities;" EPCRA Section 304 Reportable Releases: 40 CFR 355, "The List of Extremely Hazardous Substances and their Threshold Planning Quantities," Appendix A, Alphabetical Order, and Appendix B, CAS Number Order.
10. *Toxic Substances Control Act*, 15 USC §§2601–2655; 40 CFR Parts 700–799.
11. *Toxic Substances Control Act*, Chemical Substances Inventory; further information can be obtained from Industry Assistance Office, Pesticides and Toxic Substances, Environmental Protection Agency, Washington, D.C.
12. *Occupational Safety and Health Act*, 29 USC §§651–678; 29 CFR Parts 1900–1928.
13. *Occupational Safety and Health Act*, 29 CFR 1910.1000.
14. *National Environmental Policy Act*, 42 USC §§4321–4347; 40 CFR Parts 1500–1508.

General References

Subscriptions to *Federal Register* can be obtained through the Superintendent of Documents, U.S. Government Printing Office, Washington, D.C., 20402 (202) 512-1800. *Federal Register* and the *Codes of Federal Regulations* are available at those public libraries designated as U.S. Federal repositories; other government publications can be obtained from the National Technical Information Service, 5285 Port Royal Road, Springfield, Va., 22161, (703) 487-4650. A CD-ROM, including a one-year subscription with updates, is available from OSHA, which contains standards, the act, *Federal Register* notices, manuals, etc, through the U.S. Government Printing Office. The *Chemicals on Reporting Rules* list contains information on chemicals regulated by EPCRA and TSCA and is available either as paper or on diskette by calling the TSCA Hotline at (202) 554-1404.

Other subscription services

The Bureau of National Affairs, 1231 25th Street, N.W., Washington, D.C., 20037, which offers subscription services to U.S. Federal and state environmental and safety laws and regulations, both paper and on CD-ROM.
Business & Legal Reports, Inc., 39 Academy St., Madison, Conn., 06443, which offers U.S. Federal and state environmental laws and regulations, as well as such newsletters as *Environmental Manager's Compliance Advisor* and *OSHA Compliance Advisory*.

Books

Environmental Management Source Book, 1994–1995, *Environment Today*, Enterprise Communications Inc., Marietta, Ga.; annual issue as part of monthly magazine subscription.
R. L. Hoover, R. L. Hancock, K. L. Hylton, and co-workers, *Health, Safety, and Environmental Control*, Van Nostrand Reinhold Co., Inc., New York, 1989.
N. I. Sax, *Dangerous Properties of Industrial Materials*, 7th ed., Van Nostrand Reinhold Co., Inc., New York, 1989.
T. Wagner, *The Complete Handbook of Hazardous Waste Regulation*, Perry-Wagner Publishing Co., Inc., 1988.

Trade associations and other sources of information

Chemical Manufacturers Association, 1300 Wilson Blvd., Arlington, Va., 22209, (703) 741-5000, which offers information about Responsible Care (a registered trademark of CMA) and regulatory impact on the chemical industry.
Synthetic Organic Chemical Manufacturers Association, 1100 New York Ave., N.W., Washington, D.C., 20005, (202) 414-4100, which offers information about regulatory impact on the chemical industry, particularly small and batch operations.
Air & Waste Management Association, 1 Gateway Center, Pittsburgh, Pa., 15222.

Nancy R. Passow
LONZA Inc.

PHARMACEUTICALS AND COSMETICS

The U.S. Food and Drug Administration (FDA) is the primary federal scientific and regulatory government agency that monitors drug, biologic, medical devices, food, veterinary, and cosmetic products. It has been estimated that Americans spend approximately 25 cents of every dollar on FDA-regulated products (1). FDA employs approximately 9500 people, of whom chemists are the second most numerous, totaling over 1100 persons. Among the other technical personnel that compose the core of the agency are physicians, pharmacists, pharmacologists, biologists, toxicologists, microbiologists, and statisticians.

FDA's mission has traditionally been consumer protection. However, the manner in which the agency accomplishes its regulatory mission has been changing with the explosion of advances within analytical chemistry and toxicology. Because new and better compounds are continuously being developed, increased attention is being devoted to the dilemma of speeding their entry into the market while ensuring their safety and effectiveness. Scientists must not only develop new compounds, but also explain how the benefits outweigh the risks so that society can benefit from their work. Because no compound is completely safe, FDA must make an assessment of the intended use and determine whether the level of risk is acceptable. Only when FDA is able to make timely, scientifically based judgments on data will society reap the rewards of science. For example, FDA has taken the position that society is willing to accept a higher degree of risk for a life-saving device or drug than for a cosmetic or food additive. Making these judgments in a scientific and political context is the responsibility of FDA.

As a regulatory agency staffed by scientists making decisions on a daily basis, FDA's obligations are significant from both a health and safety and an economic perspective. The quality of the scientific expertise both within the FDA and relied on by the agency plays a key role in each and every FDA decision. Because FDA's regulatory foundation is tied to the basic sciences, it would be unable to perform almost any of its multiple activities without the data and information generated by the basic scientific principles and information contained in this *Encyclopedia* and other publications.

History of the FDA

American food and drug law is as old as American colonial tradition. As early as 1630, the Massachusetts Bay Colony fined Nicholas Knopp five pounds for selling a worthless concoction as a scurvy remedy. Between 1879 and 1906, some 190 separate measures affecting food and drugs were introduced in Congress. Although few of these bills passed, it was clear that with the physical and geographical redistribution of consumers and sources of consumables, urban dwellers could not protect themselves from nonobvious product defects. The processing, preserving, and pretreating of foods and the development of better drugs for common ailments led to a declining ability to protect oneself from fraud and injury. However, it was not until the passage of the Pure Food and Drug Act of 1906 that the federal government became the primary force in protecting consumers by regulating the interstate distribution of drugs and food. The 1906 Act was the first comprehensive federal law that defined and regulated adulterated and misbranded foods and drugs. From the 1906 Act, the FDA received its mission and authority to begin its role as a policing agency.

In 1938, Congress significantly modified and expanded the basic framework of the 1906 Act with the Federal Food, Drug, and Cosmetic Act (the Act). The new law extended the concept of drug regulation by requiring that prior to marketing a new drug, an application had to be submitted which provided evidence that the drug was safe. The 1938 Act, which remains the basis of later law, was enacted largely because of the improper formulation of a sulfanilamide product which caused the deaths of over 100 people. With this new Act, FDA's enforcement powers were significantly enlarged and cosmetics and medical devices came under FDA regulation for the first time (2,3).

Congress expanded FDA's authority in 1962 with the Drug Amendments of 1962, which stipulated that, before a drug could be marketed, it must be shown to be effective as well as safe. In 1976, Congress similarly expanded FDA's authority over medical devices with the Medical Device Amendments.

Other expansions of FDA's authority include the Drug Price Competition and Patent Term Restoration Act of 1984, commonly known as the 1984 Amendments or the Waxman-Hatch Act, which was passed to attain quicker marketing of safe, effective, and less expensive generic drugs; and the Safe Medical Device Amendments of 1990, which was passed to correct perceived weaknesses in the implementation of the 1976 Device Amendments. Congress further expanded FDA authority over nutrition labeling and health and nutrient content claims on food labels with the Nutrition Labeling and Education Act of 1990.

FDA Organization and Roles

The FDA is headed by the Commissioner of Food and Drugs. This position is not a Cabinet-level office but falls within the Public Health Service (PHS), a division within the U.S. Department of Health and Human Services (HHS). The post of FDA Commissioner is subject to HHS political clearance and Senate confirmation, and the Commissioner is ultimately accountable to HHS, Congress, and the President of the United States. The Commissioner has a staff to assist in policy making and several deputy commissioners to oversee operation of all the subordinate units. FDA has six regional offices within the country, each responsible for a section of the country, and 21 district offices. Persons with technical background typically work in one of FDA's chemistry laboratories or as investigators or consumer safety officers.

The FDA's approval and enforcement programs are administered by five centers organized along product lines. Although all five centers must follow the general provisions of the Act, each center is governed by its own unique and distinctive set of laws and regulations. The five centers are as follow.

- (1) Center for Drug Evaluation and Research (CDER). This center is responsible for the regulation and approval of all branded and generic human drugs, including prescription, over-the-counter, and antibiotic drugs. A drug is defined by the Act as an article intended either to be used in the diagnosis, cure, prevention, mitigation, or treatment of disease in humans or animals, or to affect the structure or any function of the body (4).
- (2) Center for Biologics Evaluation and Research (CBER). This center is responsible for the regulation and approval of all biological products intended for use in the treatment, prevention, or cure of diseases or injuries to humans. A biological product is any virus, therapeutic serum, toxin, antitoxin, vaccine, blood or blood component or derivative, or analogous product (5). It also includes products produced by biotechnology, such as interferons and erythropoietins.
- (3) Center for Devices and Radiological Health (CDRH). This center is responsible for the regulation and approval of medical devices as well as such products as x-ray machines and color television sets. A medical device is defined as an instrument, apparatus, implement, machine, contrivance, implant, *in vitro* reagent, or any related article intended either for use in the diagnosis of conditions or the diagnosis, cure, treatment, mitigation, or prevention of disease, or to affect the structure or any function of the body (6). Medical devices are distinguished from drugs in that devices do not achieve their primary intended purpose through chemical action in the body and need not be metabolized.
- (4) Center for Food Safety and Applied Nutrition (CFSAN). This center is responsible for the regulation and approval of food for human consumption, food additives, color additives, and cosmetics. Although CFSAN does not regulate meat and poultry, it does set safety and sanitation standards for supermarkets, restaurants, and other retail food establishments.
- (5) Center for Veterinary Medicine (CVM). This center is responsible for the regulation and approval of animal food and drug products. The center

also ensures that animal drugs and medicated feeds are safe and effective and that food from treated animals is safe to eat.

Independent of these centers, the FDA also has overlapping jurisdiction with several other federal agencies, including the U.S. Department of Agriculture; the Bureau of Alcohol, Tobacco, and Firearms; the Federal Trade Commission; and the Environmental Protection Agency.

Most (~90%) of the FDA's work involves enforcement of the Act. Its consumer protection function includes premarket approval and quality standards for drugs and medical devices, factory inspections, and market surveillance. FDA can combat transgressions such as the mislabeling of drugs or the adulteration of foods by issuing press releases and or warning letters, seizing products, recommending criminal prosecutions of violators to the Department of Justice, or seeking injunctions in federal courts against companies that manufacture or ship products which do not meet legal or regulatory standards necessary for consumer safety.

Because each center has its own rules and regulations, the best way to understand FDA's regulatory role is to review FDA's regulation of drug, biologic, medical device, food, veterinary, and cosmetic products separately.

Regulating Drug Products. The FDA has the authority to regulate new drugs from early laboratory research through clinical testing and market approval. In order to be approved for marketing, a new drug must be shown to be both safe and effective for its intended use (7). Safety means a low incidence of adverse reactions or insignificant side effects under adequate directions for use and warnings; it also means low potential for harm which may result from abuse under stated conditions or widespread availability. Effectiveness means a reasonable expectation that the pharmacological effect of the drug will provide clinically significant relief of the type claimed for that drug in a significant proportion of the target population when used with proper directions and warnings.

The FDA new drug approval process, which is the way most products enter the market, begins with clinical investigations. However, before clinical investigations can commence, FDA requires significant preclinical investigations involving tests on laboratory animals. These tests are to determine the nature of the chemical and to establish evidence concerning the toxicity of the substance. If, through animal studies, the drug is determined to be safe enough for human experimentation, an Investigational New Drug (IND) Application must be submitted to the Center for Drug Evaluation and Research before beginning human clinical trials. No human clinical studies may be started until 30 days after FDA has been notified.

FDA's control of clinical drug investigations is derived from the Act's prohibition of the shipment of an unapproved new drug in interstate commerce. The Act also specifically authorizes FDA to require INDs. Accordingly, FDA has implemented IND regulations that shape and control IND investigations (8). Additionally, regulations regarding the rights of human subjects, informed consent, the sale of investigational drugs, and the obligations of sponsors, monitors, investigators, and Institutional Review Boards (IRBs) have also been adopted to implement the statutory IND language (9).

Clinical investigations are broken into three phases, all of which are conducted under the oversight of an IRB to ensure that appropriate safeguards exist to protect the rights and welfare of the research subjects. The first phase of clinical investigation involves the initial administration of the drug to a small number of healthy human subjects in order to test for toxicity, drug metabolism, absorption, elimination, administration, safe dosage, and other pharmacological information. The second phase covers trials using a limited number of patients for specific disease control or for assessing diagnostic, prophylactic, or other medical use. Tests in this phase usually consist of several hundred patients and take up to two years. Less than one third of the drugs that begin the IND process typically proceed beyond this stage, usually because of safety concerns. The third phase entails large-scale clinical trials on individuals with the relevant condition. These trials can begin only if the data generated in the first two phases provide reasonable assurance that the drug is safe and effective. The third phase is intended to document safety, effectiveness, optimal dosage schedule, side effects, and directions for use in the treatment or prevention of the disease or condition. FDA typically meets with the drug firm throughout clinical trials in the third phase to identify special problems and additional testing that might be needed.

After all clinical trials are completed, the sponsor of the drug submits a New Drug Application (NDA) to FDA. The NDA consists of the IND data, manufacturing information, and data on drug stability (10). To facilitate timely review, FDA classifies all NDAs according to their therapeutic potential as compared to previously marketed drugs. Type A indicates therapeutic gain; Type B, modest therapeutic gain; Type C, little or no therapeutic gain; and Type D, both therapeutic gain and risks. Drugs proposed for use in AIDS treatments are classified as I-AA for priority review.

In order to be approved, an NDA must include data which demonstrate that the drug is both safe and effective. Each NDA is assigned to a division within CDER for consideration and administrative control, and then assigned to the appropriate therapeutic group within the division for review. The primary team of reviewers typically consists of a physician, a pharmacologist or toxicologist, and a chemist.

Other offices within CDER may become involved in the review process via consults. For example, the Office of Epidemiology and Biostatistics analyzes statistical data, the Office of Research Resources provides bioavailability reviews, and the Office of Compliance determines from the results of inspections whether the firms meet FDA's Current Good Manufacturing Practice (cGMP) regulations. Advisory committees composed of independent experts are often asked to meet and further analyze the data. Often they also advise as to what additional data and information may be needed. After FDA's review is completed, FDA issues either a Summary Basis of Approval (SBA) for the drug or a recommendation against approval. If approved, FDA releases the SBA and a summary of the safety and effectiveness data to the general public.

Regulating drug quality is a federal concern that is reflected beyond the approval process. FDA has implemented extensive regulations to ensure that drug products that are produced and marketed, as well as their chemical constituents, continue to meet high standards of quality, purity, and safety, and have the identity and strength accurately represented.

The most far-reaching program for ensuring the quality of marketed drug products is the system of cGMP regulations (11). The cGMP requirements are enforced at two stages of the development and marketing of pharmaceuticals. FDA will refuse to approve an NDA if it determines that the proposed methods, facilities, and controls are inadequate to preserve the identity, strength, quality, and purity of the drug. Once a new drug is approved, the cGMP requirements are enforced through a system of FDA inspections of manufacturing establishments.

Several classifications exist within the broad category of pharmaceuticals, each of which has its own definition and form of regulation. For example, there is a special regulatory category for drugs that are intended to treat rare diseases or conditions, ie, orphan drug products. Because the development of these drugs cannot be economically viable without some form of government assistance, Congress has passed legislation to provide incentives for drug manufacturers to develop these drugs. These incentives include a period of marketing exclusivity for approved drugs that obtain orphan status, as well as U.S. tax credits and possible direct government financial assistance.

Another distinct class of drugs are those requiring a prescription or a written order from a physician or health professional. Congress authorized FDA to determine whether a drug should be a prescription drug. Typically prescription drugs are those that (1) have habit-forming characteristics; (2) require a physician's supervision, because of toxic or other harmful effects, methods of use, or collateral measures necessary for use; or (3) are limited to prescription use under an NDA.

Drugs which are available without a prescription are readily available to consumers over-the-counter (OTC) (12). An OTC drug is low in toxicity, has low potential for harm, can be labeled for safe use without a doctor's supervision, is not habit-forming, and can be taken under easily understood conditions. The Act distinguishes between new drugs and those that are generally recognized as safe and effective. Because many OTC drugs have been marketed for years, FDA has subjected most OTC drugs to a significantly less restrictive set of regulations. This result stems from the statutory and pragmatic view that, given the agency's limited resources and the lesser hazards associated with OTC products which are generally used to alleviate symptoms rather than treat diseases, OTC products require less review. Therefore, most OTC drugs are excluded from new drug status, and no NDA has to be submitted if the active ingredient or combination of active ingredients is found to be safe and effective by the FDA as announced in final OTC drug regulations, and if the labeling of the product conforms to these OTC drug regulations.

Another subcategory of drugs are those that are reviewed under the Abbreviated New Drug Application (ANDA) process (13). These drugs are usually called generic drugs. A generic drug is one that is equivalent to a pioneer or brand-name drug but is not marketed until the brand-name drug's patent and exclusivity periods have expired. Until the 1984 Amendments, all manufacturers trying to market a new drug were required to generate their own data supporting the safety and effectiveness of their versions of the product, even if a drug with an identical active ingredient had already gone through the NDA process. The 1984 Amendments allowed generic drugs to be approved on the basis of abbreviated NDAs (ANDAs). This abbreviated approval process has the dual purpose of getting safe, effective, and less expensive generic drugs on the market, and of extending the term of patent protection to

pioneers in recognition of the need for original research by pharmaceutical companies.

Although generic drugs must meet the same standards as new drug products for identity, strength, purity, stability, adequate labeling, and bioequivalence, they need not go through the extensive clinical trials of a NDA. Instead, these generic drugs must show bioequivalence to the pioneer drug and fall into acceptable parameters set for bioavailability, which is the extent and rate at which the body absorbs the drug. By reducing the testing time, the cost of bringing the drug to market can be reduced by millions of dollars.

Finally, a different set of rules is applied to antibiotics and insulin-containing drugs (14). These categories are regulated under a monograph system mandated by statute. Thus, when a drug in these categories has been demonstrated to be safe and effective to the satisfaction of FDA, the agency promulgates a regulation of general applicability, describing in detail the required specifications of the drug. Thereafter, manufacturers meeting the standards in that regulation may obtain FDA clearance for their own product without submission of any data on safety and effectiveness, other than data demonstrating bioequivalence to the original product. Thus, later versions of a monographed antibiotic or insulin drug product are treated in a manner similar to generic drugs. Another historic distinguishing feature of this category of drugs was the requirement of batch certification. Under this requirement, the manufacturer would submit a sample from a batch of its product to FDA for testing to ensure that the batch meets the stated potency value. The batch requirement was eliminated by FDA in 1982.

Regulating Biological Products. The process for gaining FDA approval for a biological product is similar to that for a drug product. The FDA regulations require that the person or entity, eg, manufacturer, sponsoring or conducting a clinical study for the purpose of investigating a potential biological drug product's safety and effectiveness submit an IND to the Center for Biological Evaluation and Research. Clinical trials are subject to IRB review, just as drug studies. After completing the IND studies, the manufacturer submits the safety and effectiveness data generated by the studies to FDA in the form of a product license application (PLA). It is the responsibility of FDA to review the proposed labeling, the preclinical (animal and laboratory) data, the clinical (human testing) data, as well as the facilities utilized and the methodologies employed in the manufacture of the product to determine whether the product is safe and effective for its intended use. Biological products are unique in that, in addition to receiving approval of a PLA, the establishment manufacturing the biologic is subject to a prelicense inspection of the facility and the processes used to produce the potential licensed product. If both product and facility meet all standards and regulations, FDA will approve a PLA for the product and an establishment license application (ELA) for the facility (15).

Regulating Medical Devices. A person or company engaged in the manufacture, preparation, compounding, assembly, or processing of a device intended for human use must follow the regulations enforced by FDA's Center for Devices and Radiological Health. The level of FDA regulation or control is governed by the class in which the device is placed by the agency, ie, Class I, II, or III (16). Class I devices are those requiring the lowest level of regulation and are subject to general control requirements. These general controls include establishment registration; device listing; premarket notification, ie, 510(k), submission; and cGMP requirements. Class II devices are subject to special controls as well as the general control requirements. Special controls may include labeling and mandatory performance standards or other requirements. Class III devices are subject to general controls and cannot be marketed until they have an approved Premarket Approval Application (PMA) or, as a result of premarket notification (510(k)) submission, until they have been found by FDA to be substantially equivalent to preamendment devices.

Unless otherwise exempt, a firm must submit a premarket notification, also called a 510(k), to the FDA 90 days before it intends to market a device for the first time (17). The 510(k) submission must contain sufficient information to show that the device in question is substantially equivalent to a legally marketed device for a particular intended use. This notification is also required for a product when there is a change or modification to a product that may significantly affect the safety or effectiveness of the device, or when there is a significant change or modification to the intended use of the device.

Class III devices, unless they are substantially equivalent to a device already marketed without a PMA application, require formal FDA approval through the PMA process before initial sale. The PMA process is comparable to the new drug approval process (18). In both cases, safety and effectiveness data must be reviewed by FDA prior to marketing. An approved PMA application acts like a private license granted to the applicant to market a particular device. Other firms seeking to market the same type of device for the same use must also have an approved PMA.

PMA requirements differ between preamendment and post-amendment devices. Preamendment devices are those in commercial distribution before May 28, 1976; post-amendment devices are those first commercially distributed after the date. Class III post-amendment devices that are not substantially equivalent to preamendment Class III devices are considered new devices. Manufacturers of such devices are required to obtain PMA application approval before marketing these. If the post-amendment device is substantially equivalent to a preamendment device and FDA has not initiated a regulatory process specifically requiring the submission of a PMA for the device category, a 510(k) submission can be made.

To allow manufacturers to develop clinical safety and effectiveness data on devices requiring a PMA submission, FDA has implemented regulations that exempt devices intended solely for investigational use from certain provisions of the Act. This exemption is known as the Investigational Device Exemption (IDE) and allows manufacturers of devices intended solely for human investigational use to ship these products through interstate commerce (19). Like the IND regulations, the IDE regulations shape and control the investigational research. If a device is not considered to present a significant risk, an IDE submission to FDA is not necessary. If a device is considered to present a significant risk, an IDE application must be submitted to FDA for approval. In both cases, patient informed consent and IRB approval and oversight is required.

Every device manufacturer, regardless of the device class, must adhere to the requirements set forth in the device cGMP regulations (20). The essential objective of the cGMP regulations is to create a quality assurance system so that the finished device meets all the necessary specifications to maintain a high manufacturing standard. The cGMP regulations cover the methods, facilities, and controls used in preproduction design validation, manufacturing, packaging, storing, and installing medical devices. FDA monitors compliance with the cGMP regulations during its inspection of the firm's manufacturing facilities. To address the variety and complexity of devices, the cGMP regulations designate two device categories: noncritical and critical. General requirements apply to all devices, and critical devices must meet additional cGMP requirements.

Regulating Food Products. The mandate of the Center for Food Safety and Applied Nutrition (CFSAN) includes U.S. food processors, dietary supplement manufacturers, food warehouses, and cosmetic products. U.S. food processors spend \$1.4 billion annually on research and development and introduce 10,000 new grocery products each year. CFSAN monitors over 3000 food additives, thousands of pathogens, and hundreds of pesticides. In addition, the Center is responsible for handling issues involving imported food; inspecting interstate food preparations, ie, mail, planes, and boats; securing safety and sanitation standards for supermarkets, restaurants, and other retail food establishments; as well as all food labeling issues. When the 1906 Act was adopted, food adulteration and misrepresentation were rampant. Over the years, the mandate of FDA with respect to food has expanded beyond its initial role of safeguarding food against contaminants, chemical adulteration, and disease to protecting the purchaser from economic fraud, mislabeling, excessive claims, and other nonsafety offenses such as inaccurate nutrition labeling.

FDA regulates not only the finished food product, but also the ingredients that are added to food. These ingredients may be either intentionally added to food or the unintended result of materials leaching to food from product packaging. Ingredients that are intentionally added directly to food fall into two separate categories: (1) pre-1958 substances and substances generally recognized as safe (GRAS) by scientific experts, and (2) food additives (21). There are two types of food additives, these that are added directly to food and those that are not intentionally added directly to food but come into contact with food. The latter are considered indirect food additives.

An ingredient used in food prior to January 1, 1958 can be considered GRAS under the conditions of its intended use based on common use in food. FDA prior approval generally is not necessary. A post-1958 food ingredient that is generally recognized by qualified experts as safe, under the conditions of its intended use based on scientific tests, is GRAS by definition and therefore is not a food additive and does not require FDA approval prior to use.

Any substance that is not GRAS or sanctioned by use prior to 1958 (prior sanctioned) is considered a food additive. The Act prohibits the marketing of a food additive unless FDA has published a regulation that approves the intended use of the substance (22). A food additive is deemed unsafe if it is used without an approving regulation; a food is deemed adulterated if it is, bears, or contains an unapproved food additive.

To further improve the general safety standards, the Delaney Clause was included in the Food Additives Amendment of 1958. The Delaney Clause states that no food additive or color additive can be deemed safe if it has been found to induce cancer when ingested by humans or animals (23). The Clause acts as an absolute prohibition on the use of any additive found to cause cancer without any regard for whether, or to what extent, the substance is hazardous to human health. As scientific advances continue, both in the realm of food technology and analysis of previously undetected contaminants, the

zero-risk standard of the Delaney Clause will no doubt be revisited to ensure that both the goals of safety and innovation are met.

Regulating Veterinary Products. Prior to 1968, the laws surrounding veterinary products were unclear and confusing. For example, a drug for use in feeding animals, such as a penicillin product intended to help growth and prevent diseases, was classified as both a drug and a food additive because it was added to a food for animal ingestion. The burden of seeking double clearance and preparing double paperwork before being able to market a product led to strong industry support for the New Animal Drug Amendments adopted in 1968.

Animal drug controls are similar to those for human drugs. The sponsor of a new animal drug must demonstrate both safety and effectiveness of the drug for a particular intended use before a New Animal Drug Application (NADA) is approved (24). Manufacturers of generic animal drugs may submit Abbreviated New Animal Drug Applications (ANADAs), which are comparable to abbreviated new drug applications submitted by manufacturers of human generic drugs.

A distinct concern arises in the area of veterinary drugs because of the possibility that drug residues may be conveyed to humans by the food-producing animals. Therefore, drug residues and their safety in human food remain a central issue for the Center for Veterinary Medicine (CVM). Animal drugs also include those products which promotional literature claims to improve feed efficiency and increase milk production. An animal food product is regulated under the 1968 Animal Drug Amendments if it contains a drug used in feed or premixes (25).

The food additive and GRAS rules applicable to human foods generally apply to animal food ingredients. However, the Delaney clause's prohibition against carcinogenic substances in food additives was amended to permit carcinogenic chemicals to be fed to animals if the animals are not adversely affected and no residue can be found after slaughter.

FDA's medical device regulations relating to adulteration and misbranding generally apply to devices intended for use on animals. These devices, however, are exempt from the 510(k) and PMA requirements. FDA has viewed animal grooming products as being outside of its purview.

Regulating Cosmetics. Cosmetics are among the least regulated of all FDA product categories. The Act defines a cosmetic as an article, or a component of an article, intended to be used on or in the human body for "cleansing, beautifying, promoting attractiveness, or altering the appearance" of the user (26). The FDA has no statutory preapproval authority over cosmetics. FDA's enforcement mechanism against cosmetics stems from the adulteration and misbranding sections of the Act (27). A cosmetic is considered adulterated if it contains a substance which makes it harmful to users under customary conditions of use, if it contains any "filthy, putrid, or decomposed" substance, or if it is prepared under unsanitary conditions in which the product may have become contaminated.

A cosmetic is considered misbranded if the labeling is considered false or misleading. The FDA has also challenged some cosmetics as being unapproved new drugs when the product labeling suggests therapeutic or other drug value to the consumer. For example, FDA considers cosmetic products to be drugs when these products use terms such as active ingredient or claim pharmacological effects. The FDA has stated that drug–cosmetic distinction rests upon the intended use of the product, and that the administration reviews each product on a case-by-case basis. FDA has the burden of proving that a cosmetic is unsafe. This is a significant deterrent to cosmetic regulatory action because FDA must decide the amount of risk that is acceptable.

BIBLIOGRAPHY

"Regulatory Agencies" in *ECT* 3rd ed., Vol. 20, pp. 108–127, by N. R. Passow, C-E Lummus.

1. W. Schultz, testimony before the Subcommittee on Health and the Environment of the House Committee on Commerce, 104th Congress, 1st session, Washington, D.C., Feb. 2, 1995.
2. Federal Food, Drug, and Cosmetic Act of 1938 (as amended), 21 USC §§321–395 (1995).
3. 21 CFR §§1–1499 (1995).
4. Ref. 2, §321(g)(i).
5. Ref. 3, part 600.
6. Ref. 2, §321(h).
7. Ref. 2, §321(p), 355; Ref. 3, part 310.
8. Ref. 3, part 312.
9. Ref. 3, parts 312.50 and 312.56.
10. Ref. 3, part 314.
11. Ref. 3, parts 210–211.
12. Ref. 3, parts 330–358.
13. Ref. 2, §355(j).
14. Ref. 2, §§356–357; Ref. 3, parts 429–460.
15. Ref. 3, part 601.
16. Ref. 2, §306(c).
17. Ref. 2, §360(k); Ref. 3, parts 807.81–807.100.
18. Ref. 2, §360(e); Ref. 3, part 814.
19. Ref. 3, part 812.
20. Ref. 3, part 820.
21. Ref. 2, §321(s); Ref. 3, part 170.
22. Ref. 2, §348(a).
23. Ref. 2, §348(c)(3).
24. Ref. 2, §§321(w) and 360(b); Ref. 3, parts 510 and 514.
25. Ref. 2, §321(x), 360(b).
26. Ref. 2, §321(i).
27. Ref. 2, §§361–362.

General References

D. O. Beers, *Generic and Innovator Drugs: A Guide to FDA Approval Requirements*, 4th ed., Aspen Law and Business Aspen, Engelwood Cliffs, N.J., 1995.

Food and Drug Law Institute, *Seventy-Fifth Anniversary Commorative Volume of Food and Drug Law*, Food and Drug Law Institute, Washington, D.C., 1984.

Food and Drug Administration, Center for Devices and Radiological Health, *An Introduction to Medical Device Regulations*, HHS Publication No. 92-4222, U.S. Department of Health and Human Services, Public Health Service, Rockville, Md., 1992.

Food and Drug Administration, History Office, *A Guide to the Resources on the History of the Food and Drug Administration*, GPO, Washington, D.C, 1995.

J. E. Foulke, *Cosmetic Ingredients*, HHS Publication No. 93-5013, Department of Health and Human Services, GPO, Washington, D.C., 1993.

Decoding the Cosmetic Label, HHS Publication No. 95-5016, Department of Health and Human Services, GPO, Washington, D.C., 1995.

W. F. Janssen, *The U.S. Food and Drug Law: How It Came, How It Works*, HHS Publication No. 92-1054, Department of Health and Human Services, GPO, Washington, D.C., 1992.

J. T. O'Reilly, *Food and Drug Administration*, 2 vols, 2nd ed., Shepard's McGraw-Hill, Colorado Springs, Colo., 1993.

D. Stehlin, *Cosmetic Safety More Complex Than At First Blush*, HHS Publication No. 93-5012, Department of Health and Human Services, GPO, Washington, D.C., 1993.

Gary L. Yingling
Suzan Onel
McKenna & Cuneo, LLP

POWER GENERATION

Entities involved in the generation of electric power in the United States are subject to regulation by numerous state and federal agencies. As with other industrial processes, there are a plethora of environmental regulations and laws in place, designed to ensure that the generation and delivery of electric power have minimal impact on air, land, and water resources. These regulations and laws are also designed to protect the welfare of the ecosystem, public health, and overall quality of life.

In addition to environmental regulation, there are many laws and regulations in place that are designed to ensure fair business practices, a stable energy supply, and consumer protection. However, some of the key regulatory issues impacting the industry are likely to change by the year 2000 as environmental regulation continues to become more stringent and business-practice regulation undergoes a significant transition when more competition is introduced into the power market.

The Federal Energy Regulatory Commission (FERC), an independent agency of the U.S. government, has regulatory oversight of the country's electric utilities, natural gas pipeline transportation system, oil pipeline transportation system, and most U.S. hydroelectric generation facilities (1). More specifically, in the power generation sector, FERC administers laws and regulations involving critical energy issues, including transmission and wholesale sales of electric energy in interstate commerce; licensing and inspection of private, municipal, and state hydroelectric projects; and oversight of related environmental matters.

FERC was created in October 1977 as a result of the passage of the Department of Energy Organization Act. At that time, the 57-year-old Federal Power Commission (FPC) was eliminated and FERC took over many of the FPC's responsibilities.

History of Electric Utility Business Regulation

The history of regulation (1–4) in the United States sheds light on the current situation. Thomas Edison's Pearl Street Generating Station went into business in 1882, marking the start of the U.S. electric utility industry. It was the first U.S. power plant and served 85 customers, lighting 400 lights in downtown Manhattan near New York City's financial district. As the industry developed over the following decade, it became apparent that building more than one utility system, including power plants, substations, and transmission and distribution networks, in a given area was not economical. It was also apparent that there were economies of scale associated with using large central generating stations to serve a large customer base. There were also many niche markets for smaller power plants which could be sited at or near industrial loads. By 1900, such plants represented more than half of the country's installed generating capacity. However, electric utility plants accounted for more than 85% of the country's installed generating capacity by 1950, and more than 90% by 1993.

Utility systems were extremely capital-intensive and investors needed to know that a stable customer base could be secured. Thus, electric companies typically secured rights to franchise territories. At the turn of the century, franchises were usually granted by city councils and local politicians. Because the franchises were granted and controlled by these partisan groups, they were also subject to the instability associated with ever-changing leaderships. Graft and favoritism associated with franchise awards were not uncommon. In some areas, municipal-owned systems were developed to help overcome the problems associated with granting service territories and to help ensure a stable, low cost power supply.

By 1907, both utilities and consumers were speaking out in favor of establishing state regulation that would put the power of utility oversight in the hands of nonpartisan agencies. In that year, three states formed regulatory agencies; by 1916, 33 states had regulatory agencies. These agencies treated utilities as natural monopolies. Because only one utility could economically serve a given area, state regulators would set the electric companies' rates to ensure that consumers had an affordable power supply. The rates were also designed to ensure that the utility received a fair return on investment, a factor that stabilized the industry and enabled utilities to secure the long-term financing that was a necessity for growth.

Between 1900 and 1930, the electric power industry grew rapidly in the United States. As electricity became less expensive and widely available, new uses for electric power were continually developed. During this time period, the principal players in the utility business formed large holding companies that controlled numerous subsidiaries, including electric power equipment suppliers, engineering and construction firms involved with utility system design and construction, project finance companies, and system plant operating companies. Some of the holding companies were used to shift profits away from the state-regulated utilities that operated plants and sold power. For example, the unregulated holding company would overcharge its utility subsidiary for arranging project financing and providing engineering and construction services.

The excesses of several of the holding companies attracted the attention of government leaders and, in 1928, the Federal Trade Commission (FTC) began investigating the holding companies. This led to the Public Utility Act of 1935, which amended the earlier Federal Water Power Act of 1920 and expanded the duties of the Federal Power Commission under the newly amended Federal Power Act to include regulation of electric utility wholesale and transmission rates and transactions. It also required interstate holding companies to register with the Securities and Exchange Commission (SEC) and to operate according to the SEC's rules, and that certain holding companies be broken up by the SEC.

The next significant change in federal regulations affecting the entire U.S. power industry did not occur until after the FPC was replaced by the Federal Energy Regulatory Commission in the 1970s. In the wake of the energy crisis, Congress passed the Public Utility Regulatory Policies Act of 1978 (PURPA) as part of the National Energy Act of 1978. PURPA aimed to increase conservation of energy and improve the efficiency of the use of facilities and resources by electric utilities. It called for state regulatory authorities to encourage conservation and utility efficiency and to provide for equitable rates. Some of PURPA's provisions were designed to encourage the development of cogeneration and small power production facilities, typically located at industrial facilities, by loosening the economic, regulatory, and institutional barriers that discouraged cogeneration and the use of renewable energy resources.

Obstacles previously faced by nonutility generators (NUGs) before PURPA included (1) utilities typically had no incentive to pay NUGs attractive rates for their power output because the unregulated NUGs usually represented competition for utilities, particularly for large industrial loads; (2) some utilities charged high rates for backup power services to NUGs; and (3) a NUG wishing to sell power into an electric utility's system risked being considered a utility itself and could be subject to extensive federal and state regulation.

In 1986, Congress passed the Electric Consumers Protection Act (ECPA), which set the foundation for more efficient relicensing of hydropower facilities, enhanced competition among applicants for relicensing, and ensured a balance between environmental interests and the need for hydropower development.

Comprehensive changes to power industry regulation are underway in the 1990s as a result of Congress' passage of the Energy Policy Act (EPACT) of 1992, which includes provisions designed to foster competition through regulatory reform in both the oil pipeline and electric utility industries. EPACT also amends the 1935 Public Utility Act, which was designed to discourage holding companies structured to hamper effective state regulation. Although its implementation is still being negotiated (1996) by federal and state government as well as industry groups, EPACT is likely to impact the power industry in three ways. First, EPACT allows FERC to order upon application wholesale, but not retail, transmission line access. Again, EPACT thus aims to promote wholesale power competition. Second, EPACT establishes a program for providing federal support on a competitive basis for renewable energy technologies. It also expands a program to promote the export of renewable technologies to emerging markets in developing countries. Third, a class of

power generation entities known as exempt wholesale generator (EWG) has been created. EWGs, which may be owned by either utilities or nonutilities, are exempt from the corporate organizational restrictions of the 1935 PUHCA. EPACT thus aims to remove obstacles to wholesale power competition that are included in PUHCA.

Federal Energy Regulatory Commission

Organization. FERC has five members who are appointed by the President of the United States, with the advice and consent of the U.S. Senate, to five-year staggered terms. Each commissioner has an equal vote on regulatory matters. No more than three commissioners may belong to the same political party. One member is designated by the President to serve as chairperson and is the commission's administrative head. The commission generally meets publicly twice per month to discuss agency business, including setting of industrywide rules and consideration of license applications, rate filings, and other matters submitted by regulatory entities.

FERC regulates the sale of electric energy at wholesale and the transmission of electric energy in interstate commerce by public utilities under the Federal Power Act, but the retail sale of electricity, ie, sales to end-use customers such as homeowners and businesses by public utilities, is generally regulated by state public utility commissions pursuant to state laws. Sales of electricity for resale (FERC's regulatory domain) represent approximately 25% of total electric sales by investor-owned utilities in the United States. When a public utility files for rate changes or modifications to its terms or conditions of electric service, the commission issues a public notice, listed in the *Federal Register*, soliciting comments, protests, and interventions. Approximately 85% of the Commission's rate filings are processed by FERC staff members through delegated authority. The commission itself handles only major rate increase requests and contested applications, and, pursuant to the Federal Power Act, acts on these requests and applications by either accepting all or part of an application, rejecting all or part of an application, or suspending the effectiveness of the application and ordering a hearing and investigation.

The commission authorizes the sale, lease, or disposition of certain utility facilities, mergers or consolidation of such facilities (mergers), and certain security issuances (financial instruments) by public utilities. Under the Federal Power Act, FERC also reviews the relationships of officers and boards of directors of utilities to other utilities, firms supplying electrical equipment to the utilities, or firms underwriting securities. Finally, under separate statutory authority, FERC reviews rates charged by the federal power marketing administrations and, pursuant to PURPA and EPACT, certifies small power production and cogeneration facilities, ie, entities defined by PURPA, as well as exempt wholesale generators, ie, entities defined by EPACT.

In the hydroelectric area, pursuant to the Federal Power Act as amended, FERC's responsibilities include project licensing, dam safety, project compliance, investigation and assessment of headwater benefits, environmental issues, and interagency coordination. FERC engineers perform periodic safety inspections of hydroelectric power facilities. These inspections emphasize the evaluation of structural integrity of hydro plant structures. The commission's licensing costs are offset by annual charges to the license holders. Finally, FERC determines charges for the use of certain federal lands and dams as well as Native American reservations.

Policies and Rulemaking. In March of 1995, FERC proposed sweeping changes in the electric utility industry. At that time, FERC declared that "monopoly control of the nation's transmission system is the single greatest impediment to wholesale competition" for electric power sales (5), and released a Notice of Proposed Rulemaking (NOPR) seeking comments on FERC's proposals to encourage a more fully competitive wholesale electric power market. The proposed rule would require utilities under FERC jurisdiction that own or control transmission facilities in interstate commerce to file tariffs under which they will provide service to third parties seeking access to the system, ie, for transport of electric power. The rule also requires utilities under FERC jurisdiction to offer to eligible customers transmission service comparable to the service they provide themselves, and to take service under the same tariffs for their own wholesale sales and purchases of electric energy.

The proposed rule includes two *pro forma* tariffs, one for network service and one for point-to-point service, setting out the minimum terms and conditions which the utility must offer. These terms and conditions specifically cite access to ancillary services, including scheduling and dispatching, load following, imbalance resolution, reactive power support, loss compensation, and system protection. In addition, the proposal includes requirements for utilities to enlarge their transmission capacity and expand ancillary services as needed for requested transmission access. The proposed rule also states that a public utility must obtain transmission services, including ancillary services, for all of its new wholesale sales and purchases of electric energy under the same tariff with which it offers such services to others, and a transmission owner's tariff must include separately stated rates for transmission and ancillary service components. Also, a public utility must rely on the same electronic information network that its customers use to obtain transmission information about its system when buying or selling power.

FERC's March 1995 NOPR has been the subject of much controversy and debate. State regulatory commissions are struggling to determine how to regulate a restructured power industry. Electric utilities are lobbying to ensure that they are fully reimbursed for any stranded assets, ie, assets rendered uneconomic, as a result of FERC's proposed rules. For instance, a large industrial customer may choose to purchase power from an entity other than its traditional utility provider, much like choosing a different long-distance telephone service provider. As a result of the loss of these large customers, a utility may be saddled with large generation facilities that are no longer economic to operate. The utilities maintain that they should be compensated for losses in such cases because the facilities were financed and built under the premise that a guaranteed return on investment could be expected because of the existing regulatory environment.

Entities involved in long-term contracts with electric utilities, such as fuel suppliers and NUGs selling power to utilities, also have concerns that some utilities or industrial customers will not be able to honor their contracts under the new, more competitive system. Finally, some utilities are concerned that they will not be adequately reimbursed for opening up their transmission systems to competitors. The potential competitors in turn are concerned that utilities will not provide unbiased access to their transmission systems if the utilities themselves are also in business of marketing power. There has also been some debate regarding which transmission facilities are eligible for open access. This is because some facilities are considered local distribution systems by utilities, which feel they should not be opened to competitors.

FERC believes that 137 utilities will be required to open up their transmission networks. As of mid-1996, the NOPR is still being debated. However, several states have adopted preliminary plans to open up the transmission networks of their regulated utilities by the year 2000 or shortly thereafter.

Environmental Regulations

The electric power industry is subject to many of the same environmental regulations and laws affecting other process industries which have the potential to generate waste streams (see REGULATORY AGENCIES). The main agencies involved in the environmental regulation of power generation facilities are the Environmental Protection Agency (EPA), the Nuclear Regulatory Commission (NRC), and the Occupational Safety and Health Administration (OSHA). Because power generation can potentially impact all aspects of the environment, there are volumes of regulations dictating what emissions levels are allowable, including airborne pollutants, solid and liquid wastes, noise, and inert discharges such as water. There are also many regulations that impact where a power plant can be sited and how it must be designed to minimize its impact on the surrounding environment. In many cases, federal agencies set national standards which individual states must implement. In addition, state-specific regulations for pollution control are often more stringent than federal regulations. For specific information about environmental regulations affecting a given geographic area, readers should consult the local office of their state department of environmental protection, the appropriate regional environmental commission, or the local EPA office.

The Clean Air Act of 1970 was the first federal environmental legislation to affect significantly the electric power industry. It was in that year that the Environmental Protection Agency was empowered to set enforceable air quality standards. However, Congress had previously passed the Clean Air Act (CAA) of 1963 in response to growing concerns about airborne power plant exhaust emissions of sulfur dioxide, SO₂, which is a precursor of acid rain containing sulfuric acid, and oxides of nitrogen, NO_x, which is a principal contributor to low level ozone, ie, smog, and a lesser contributor to acid rain containing nitric acid (4). Prior to 1970, SO₂ regulations were focused on reducing ground-level SO₂ emissions to prevent associated impacts to human health. This is one of the reasons for using tall exhaust gas stacks, which help disperse airborne emissions high into the atmosphere to minimize terrestrial impacts.

In 1971, EPA established New Source Performance Standards (NSPS), which required coal-fired utility boilers built after August 17, 1971 to emit no

more than 0.52 kg SO₂ 10³ kJ (1.2 lb of SO₂ 10³ Btu) of fuel heating value input. Requirements for NO_x ranged from 0.086–0.34 kg 10³ kJ (0.2–0.8 lb 10³ Btu) of fuel input, depending on the type of fuel burned and the combustion device used.

In 1977, Congress amended the CAA to require states to set limits on existing sources in regions not attaining goals established in the CAA of 1970. In 1979, EPA established the Revised New Source Performance Standards (RNSPS), which retained the 1971 NSPS of 0.52 kg 10³ kJ for SO₂ emissions, but required SO₂ emissions from all new or modified boilers (post-1978) to be reduced by at least 90% , unless emissions could be cut to 0.26 kg 10³ kJ (0.6 lb 10³ Btu) at less than 90% removal. In those cases, 70–90% removal are permitted, depending on the sulfur content of the coal. RNSPS for NO_x are more complex. As with NSPS levels, RNSPS NO_x requirements varied from 0.086–0.34 kg 10³ kJ, depending on fuel content and the combustion technology used. However, RNSPS for NO_x delineate the combustion devices into more categories.

Clean Air Act Amendments of 1990. The primary aspects of the Clean Air Act Amendments (CAAA) of 1990, which impact electricity producers, include a 10 million ton (short ton) reduction in national SO₂ emissions and a two million ton reduction in NO_x emissions from 1980 levels (Tables 1 and 2). The reduction in SO₂ is occurring in two phases. Phase I began in 1995 and Phase II begins in 2000. The CAA established an innovative program whereby companies could purchase and trade allowances for SO₂ and NO_x emissions as one means of meeting the new limits. In the power generation sector, the CAAA focus mainly on electric-utility-owned generation capacity.

Table 1. Estimated Emissions from Fossil-Fueled Steam-Electric Generating Units at U.S. Electric Utilities, 10³ t^a

Emission ^b	1989	1990	1991	1992	1993
sulfur dioxide, SO ₂	14,297	13,943	13,619	13,318	13,093
nitrogen oxides, NO _x	5,403	5,263	5,263	5,147	5,309
carbon dioxide, CO ₂	1,739,652	1,700,232	1,704,483	1,690,441	1,765,653

^a Refs. 6 and 7.
^b Estimates for 1993 are preliminary; for prior years are final. Estimates are for steam-electric plants >10 MW, based on fuel consumption data.

Table 2. 1993 Estimated Emissions from Fossil-Fueled Steam-Electric Generating Units at Electric Utilities by Census Division, 10³ t^{a, b}

Census division state	Coal			Petroleum			Gas ^c		Other ^d		
	Sulfur dioxide	Nitrogen oxides	Carbon dioxide	Sulfur dioxide	Nitrogen oxides ^e	Carbon dioxide	Nitrogen oxides	Carbon dioxide	Sulfur dioxide ^f	Nitrogen oxides ^g	Carbon dioxide ^h
New England ⁱ	119	47	15,286	107	21	13,551	6	1,681	1	1	969
Middle Atlantic ^j	1,383	347	130,307	87	28	19,594	31	12,893			45
East North Central ^k	4,661	1,453	392,429	8	3	2,230	6	2,138	1	1	785
West North Central ^l	1,016	649	188,637	2		315	7	1,882	1	1	1,098
South Atlantic ^m	3,050	941	334,732	297	65	37,523	32	9,853	1		292
East South Central ⁿ	2,297	677	219,866	53	6	3,292	5	1,756	0	0	0
West South Central ^o	783	622	215,416	3	1	729	254	85,476	0	0	0
Mountain ^p	452	508	201,553	1	1	403	11	3,880	0		
Pacific contiguous ^q	83	45	13,447	4	3	1,738	71	27,024			690
Pacific noncontiguous ^r	1	0	0	20	9	4,754	0	0	0	0	0
Total	13,844	5,288	1,711,673	583	136	84,129	424	146,584	4	4	3,880

^a Refs. 6 and 7.
^b Estimates for 1993 are preliminary. Estimates are for steam-electric plants >10 MW, based on fuel consumption data.
^c Except for Pacific noncontiguous, which has a value of zero, sulfur dioxide emissions for all other census divisions are <0.5; total sulfur dioxide value 1.
^d Includes light oil, methane, coal–oil mixture, propane gas, blast furnace gas, wood, and refuse.
^e Emission value is <0.5 for West North Central.
^f Emission value is <0.5 for Middle Atlantic and Pacific contiguous.
^g Emission value is <0.5 for Middle Atlantic, South Atlantic, Mountain, and Pacific contiguous.
^h Emission value is <0.5 for Mountain.
ⁱ Includes Conn., Maine, Mass., N.H., R.I., and Vt.
^j Includes N.J., N.Y., and Pa.
^k Includes Ill., Ind., Mich., Ohio, and Wis.
^l Includes Iowa, Kans., Minn., Mo., Nebr., N.D., and S.D.
^m Includes Del., D.C., Fla., Ga., Md., N.C., S.C., Va., and W. Va.
ⁿ Includes Ala., Ky., Miss., and Tenn.
^o Includes Ark., La., Okla., and Tex.
^p Includes Ariz., Colo., Idaho, Mont., Nev., N. Mex., Utah, and Wyo.
^q Includes Calif., Oreg., and Wash.
^r Includes Alaska and Hawaii.

Sulfur Dioxide Control. Title 4 of the CAAA has the primary goal of reducing annual SO₂ emissions by 10 million tons below 1980 levels. A national cap on SO₂ emissions of 15 million tons per year has been imposed, including 8.95 million tons yr for utilities and 5.6 million tons yr for industrial sources (8). To achieve such reductions, a two-phase compliance plan was established. Phase I, effective January 1, 1995, impacts 110 large coal-fired plants, comprising 261 separate units and representing approximately 89,000 MW of capacity or over 10% of the country's installed capacity. These facilities are located in 21 eastern and midwestern states. Owners of these units were given the option of meeting the new requirements by any of the following means: installing SO₂ emissions control devices such as wet flue gas desulfurization systems; switching to lower sulfur fuel; applying to EPA to

substitute control of other units to attain the required emissions reductions; and obtaining additional emissions allowances by overcontrolling emissions from one or more units.

Utilities that reduce emissions below the number of allowances they hold may trade emissions credits on the open market. Owners of plants affected by Phase I regulations can also petition the EPA for a two-year extension for meeting Phase I emissions if they have selected a control option capable of reducing SO₂ emissions by 90% or more, such as is capable by flue-gas desulfurization. Owners of these units can receive bonus allowances for 1997–1999 if they have operated at SO₂ emissions below 0.52 kg 10 kJ (1.2 lb 10 Btu) of fuel heating value input.

Phase II of the Title 4 of the CAAA begins in the year 2000, which aims to tighten emissions down on all existing utility-owned oil- and coal-fired units larger than 25 MW and all new utility plants. Starting in 2000, units that have been in operation since before 1985 will be required to lower emissions to the lesser of either 0.52 kg 10 kJ for SO₂ emissions or whatever their existing NSPS requirement was. However, special allowances will be provided for certain units that were previously fueled by oil or gas, that started operation between 1985 and 1996, that operated at lower capacity factors in the mid-1980s, that are part of small utility systems or located in states having historically low emissions rates, that are part of certain municipal systems, or that are located in Florida or 10 other listed states (9).

Any units that started operation after December 31, 1995 not only received no allowances, but were required to purchase them to offset emissions. Certain units may be eligible for a four-year extension for meeting Phase II SO₂ emissions limits. To be eligible, a plan must be in place before January 1998 to repower the affected facility using one of a listed group of clean coal repowering technologies. In this case, repowering refers to the replacement of a plant's existing boiler with some other advanced, low emission combustion system, such as a fluidized-bed boiler or an integrated gasification combined-cycle unit (see POWER GENERATION).

CAAA Impact on Nonutility Power Producers. The SO₂ and NO_x regulations being implemented as part of the CAAA of 1990 primarily target electric utility power plants. However, under Phase II of the CAAA, nonutility power producers will be required to acquire emissions allowances for any SO₂ being emitted from new facilities. Although industrial emitters of SO₂ and NO_x are not directly affected, the EPA did undertake a study to estimate what contribution industrial producers have on annual estimated SO₂ production in the United States (10). The report found that annual industrial SO₂ emissions would remain below the predetermined critical limit of 5.6 × 10⁶ tons yr until at least 2015 (10). Thus, the agency recommended no new controls for industrial SO₂ emissions at this time.

Nonutility generators can be affected by parts of the CAAA's NO_x reduction measures. This impact will vary, depending on what state a nonutility generator is located in, what technology is employed, and what fuel is burned. Title 1 of the 1990 CAAA targets NO_x reductions in approximately 100 different geographic areas of the United States where national ambient air quality standards for ground-level ozone are being exceeded.

NO_x Control. NO_x control limitations are described in both Title 1 and Title 4 of the CAAA of 1990. Title 4 requirements affect only coal-fired boilers and take effect at the same time that the boilers are impacted by CAAA SO₂ requirements. As of 1996, EPA had established Title 4 NO_x limits only for tangentially fired and wall-fired, dry-bottom boilers that would be impacted by Phase I of the CAAA SO₂ regulations (Title 4). Limits of 0.22 kg 10 kJ (0.5 lb 10 Btu) and 0.19 kg 10 kJ (0.45 lb 10 Btu) have been set for wall-fired and tangentially fired units, respectively. The EPA based these levels on what was achievable using low NO_x burners. However, plants can employ a number of different front- or back-end emissions controls, including a combination of options, to achieve these levels. EPA plans to announce Title 4 NO_x requirements for 300 additional boilers by late 1996 or early 1997.

Title 1 of the CAAA of 1990 focuses on utility units located in ozone nonattainment regions of the United States, which include approximately 100 different areas where national standards for ground-level ozone (O₃) are exceeded. More than 900 utility boilers are located within the nonattainment zones, including 90 units impacted by Phase I of Title 4. These boilers represent approximately 400,000 MW of installed capacity and 60% of the country's utility boilers.

The impacts of Title 1 of the CAAA vary, depending on how serious the nonattainment problem is in a given area. In areas where the degree of nonattainment is more severe, local or regional NO_x requirements are typically more stringent.

In each region, states are developing implementation plans to achieve compliance for ground-level ozone. EPA has provided a framework for guiding the states. For example, starting in November 1992, all major new sources of NO_x were required to meet permit limits representing lowest achievable emissions rates (LAER). In addition, NO_x emissions from any new units must be offset, to varying degrees, by reductions in NO_x made elsewhere. The definition of a major source varies, depending on the severity of the ozone problem. For example, in the Los Angeles area, which is considered one of the country's most extreme areas for ozone nonattainment and in which O₃ levels in the air exceed 0.28 ppm, a major source is considered any unit producing NO_x at a rate of 10 tons yr or more. In contrast, in moderate nonattainment areas, such as parts of Arizona and southeast Florida, where O₃ levels range from 0.139–0.160 ppm, a major new source is a producer adding more than 100 tons yr.

Starting in May 1995, all existing major NO_x sources located in ozone noncompliance areas were required to employ reasonably available control technologies (RACT) for NO_x control. Under EPA regulatory supervision, states are responsible for determining what specific technologies are considered RACT for a given type of plant and what specific emissions regulations limits should be. In many cases, states have adopted regulations that are tighter than those suggested by the EPA. As of 1996, information on state-specific RACT programs is still being compiled. However, a variety of technologies have been applied on utility boilers to meet the first phase of NO_x reductions, including low NO_x burner retrofits, flue-gas recirculation, natural gas cofiring, combustion air modifications, and selective noncatalytic reduction (SNCR). On gas turbine-based plants, NO_x compliance options have included low NO_x burner retrofits, water or steam injection, SNCR, and selective catalytic reduction (SCR).

Because of the necessity to comply with national standards for ground-level ozone, some states are planning another phase of more stringent NO_x emissions limits which may take place in the early 2000s. These additional post-RACT reductions may affect plants of all sizes and types, but are likely to focus on major sources. The deadline for compliance in the most extreme areas is 2010. For severe nonattainment areas (O₃ levels 0.181–0.280 ppm), including many coastal areas in the Northeast, from northern Virginia to southern Maine, compliance must be achieved by November 2005 to November 2007. Serious ozone nonattainment areas (O₃ levels 0.161–0.180 ppm) are expected to be in compliance by November 1999. Moderate noncompliance areas must comply by November 1996.

Hazardous Air Pollutants. Title 3 of the CAAA of 1990 addresses the release of hazardous air pollutants (HAPs) by requiring both the identification of major stationary sources and area source categories for 189 toxic chemicals and the promulgation of control standards. Major sources of air toxics, also referred to as HAPs, include any stationary source or group of sources emitting 10 or more tons yr of any single listed toxic chemical or 25 tons yr of a combination of any listed toxic. Area sources of HAPs include smaller plants that emit less than the 10 or 20 tons yr thresholds. The major sources of HAPs are typically industrial facilities. However, Title 3 requires the EPA to study potential health affects associated with emissions of HAPs from electric utility boilers (11).

EPA is also required to recommend maximum achievable control technologies (MACT) for reducing HAPs from new and existing major and area sources to be achieved by 2000. After that time, additional control measures may also be considered. As of 1996, EPA has recommended no new HAP for power boilers. However, both utility and nonutility power generators are keeping a close watch out for possible future regulations for certain HAPs, including mercury and hydrogen chloride.

Other Emissions. Title 3 of the CAAA also impacts power plant particulate matter (ash) emissions. In June of 1994, the EPA actually relaxed its standards for emissions of particulate smaller than 10 micrometers (PM₁₀). This revision was in response to the EPA's mandate to review health-based pollution standards every five years (12). However, it is uncertain as of this writing (1996) if states will indeed implement less stringent regulations for PM₁₀ emissions.

The CAAA also established a requirement that major new or modified sources of PM₁₀ or carbon monoxide (CO) located in ozone nonattainment

areas must obtain emissions offsets from areas of equal or worse nonattainment before construction of such facilities is approved. There are already regulations in place, including the Clean Drinking Water Act and the Resource Conservation and Recovery Act, which limit the discharge of toxics in solid and liquid discharges from power plants. Some industry observers believe that the increased use of natural gas as a power plant fuel may help to reduce overall electric utility emissions of fine particulates, HAPs, and carbon dioxide. Increasing natural gas utilization through cofiring or fuel switching is one of the options that utilities are using to comply with NO_x and SO₂ emissions requirements of the CAAA.

Continuous Emissions Monitoring. A key aspect of the new CAAA is the requirement that plants prove their continued compliance to new emissions limits by installing continuous emissions monitoring systems (CEMs). The CAAA imposes new requirements for monitoring NO_x, SO₂, and CO₂ levels in a plant's exhaust gas stream. Affected plants typically must gather data from stack monitoring systems, gas analyzers, and the plant's data acquisition system and provide the data in a format approved by the EPA and state regulators. CEM systems must be in place by November 1993 for boilers affected by Phase I of the CAAA, and by January 1995 for plants impacted by Phase II.

Nuclear Regulatory Commission

The U.S. Nuclear Regulatory Commission (NRC) was formed in 1975, following the Energy Reorganization Act of 1974, to regulate the various commercial and institutional uses of nuclear energy, including nuclear power plants. The agency succeeded the Atomic Energy Commission (AEC), which was formed under the Atomic Energy Act of 1946 (13).

In 1954, Congress passed additional atomic energy legislation that redefined the U.S. atomic energy program by ending the federal government's secrecy concerning certain atomic energy information. The AEC was instructed by Congress to encourage widespread participation in the development and utilization of atomic energy for peaceful purposes. Following the 1954 Act, the AEC played three principal roles: (1) managing the U.S. atomic weapons program; (2) promoting the private use of atomic energy for peaceful purposes, such as power generation; and (3) protecting public health and safety from the potential hazards of nuclear-related programs. At that time, the AEC began establishing guidelines for nuclear power plant licensing as well as radiation protection guidelines.

When the NRC, headquartered in Rockville, Maryland, took over the responsibilities of the AEC in 1974, many of the AEC's research and development functions, particularly many covering new technology development and nuclear weapons production, were assumed by the U.S. Department of Energy. However, the NRC has maintained some research and developmental capabilities which are handled by the NRC's Office of Nuclear Regulatory Research.

In its responsibility to protect public health and safety, the NRC has three principal regulatory functions: (1) establishing standards and regulations; (2) issuing licenses for nuclear facilities and users of nuclear materials; and (3) inspecting facilities and users of nuclear materials to ensure compliance with NRC's requirements. These regulatory functions apply to both nuclear power plants and to other uses of nuclear materials, such as nuclear medicine programs at hospitals, academic and research activities, and manufacturing of products containing radioactive materials, eg, radiation sources found inside of smoke detectors or luminous watch dials.

The NRC issues licenses for the facilities noted and the operators of those facilities. Licenses may also be issued by individual state governments under NRC-approved regulatory programs. There are more than 8500 such licenses under the NRC's jurisdiction and approximately 15,000 under the jurisdiction of Agreement States, which regulate certain radioactive materials under agreements with the NRC. As of 1996, there are 109 licensed commercial nuclear power reactors in the United States, located at 71 sites in 33 states (see NUCLEAR REACTORS). However, several of these facilities are only partially constructed and further construction has been deferred. There are more than 5300 licensed nuclear power plant operators in the United States, each licensed for a specific reactor. Every operator must be requalified before renewal of a six-year license (14,15).

The NRC has numerous offices and special committees, including more than 3000 staff members, which serve to advise and support the five commissioners appointed to staggered five-year terms by the President. The NRC's activities have been described in the *Nuclear Regulatory Information Digest*, 1993 (14).

The NRC relies primarily on reactor and facility inspections as the basis for licensee compliance with NRC regulations. Thus, the NRC personnel conduct from 10–30 routine inspections each year at every operating nuclear plant, depending on the activities underway at individual plants and problems which may occur. Inspection specialists, based in regional offices, review plant security, emergency planning, radiation protection, environmental monitoring, periodic testing of plant equipment and systems, fire protection, construction activities, and a number of other specialized areas.

NRC inspection teams issue inspection reports and provide these to the facilities, as well as designated public libraries located near operating plants and the NRC's Public Document Room for public disclosure. When inspections disclose violations of NRC requirements, the agency immediately issues a Notice of Violation, which requires the licensee to correct the problem and prevent recurrence. The licensee is given a deadline for planning and undertaking an acceptable response. For certain more serious or repetitive violations, the NRC may fine utilities or other licensees up to \$100,000 per day for each violation. If inspection teams determine that a facility is operating in an unsafe manner, the NRC may issue orders requiring the shutdown of a facility or the modification of previous operating procedures.

Nuclear Waste. NRC defines high level radioactive waste to include (1) irradiated (spent) reactor fuel; (2) liquid waste resulting from the operation of the first cycle solvent extraction system, and the concentrated wastes from subsequent extraction cycles, in a facility for reprocessing irradiated reactor fuel; and (3) solids into which such liquid wastes have been converted. Approximately 23,000 metric tons of spent nuclear fuel has been stored at commercial nuclear reactors as of 1991. This amount is expected to double by the year 2001.

As of 1996, the bulk of spent fuel from nuclear power plants has been stored in specially designed water-filled holding pools at the reactor site. Some spent fuel is also stored at two off-site storage facilities, one in Illinois and one in upper New York State (which no longer accepts waste). The water in these holding pools acts as a shield preventing release of radioactivity. At some plants, waste is beginning to be stored dry in heavily shielded, NRC-approved iron, steel, or concrete casks located on concrete pads.

Approximately 25–30% of a reactor's fuel is removed and replaced during planned refueling outages, which normally occur every 12 to 18 months. Spent fuel is highly radioactive because it contains by-products from nuclear fission created during reactor operation. A characteristic of these radioactive materials is that they gradually decay, losing their radioactive properties at a set rate. Each radioactive component has a different rate of decay known as its half-life, which is the time it takes for a material to lose half of its radioactivity. The radioactive components in spent nuclear fuel include cobalt-60 (5-yr half-life), cesium-137 (30-yr half-life), and plutonium-239 (24,400-yr half-life).

Although most spent fuel is stored on-site in the 1990s, a number of facilities are expected to run out of space within the next two decades. Some waste is still transported to the only active off-site storage facility located in Illinois. In addition, the federal government is trying to site a large centralized storage facility capable of handling all of the country's high level waste.

The NRC has developed special procedures for the handling, transportation, and storage of nuclear fuel because radioactivity can be a health hazard if not properly shielded. Spent fuel is typically transported by rail or truck in heavily shielded (Type B), sealed, thick metal shipping containers designed to withstand possible accidents, such as derailments or collisions, which may occur during transport. The NRC certifies that each shipping container meets federal requirements. The U.S. Department of Transportation sets the rules for transportation.

The NRC also imposes special security requirements for spent fuel shipments and transport of highly enriched uranium or plutonium materials that can be used in the manufacture of nuclear weapons. These security measures include route evaluation, escort personnel and vehicles, communications capabilities, and emergency plans. State governments are notified in advance of any planned shipment within their state of spent fuel, or any other radioactive materials requiring shipment in accident-proof, Type B containers.

In 1980, Congress determined that each state should be responsible for ensuring the proper handling and disposal of commercial low level nuclear wastes generated in their states. Regional disposal sites have also been established at Barnwell, South Carolina, and Ward Valley, California. These wastes are handled by licensed disposal facilities where they are packaged, placed in burial trenches, and covered with soil. Less than half of the low level nuclear waste produced annually in the United States comes from nuclear power plants. Low level nuclear power plant wastes include contaminated equipment,

filters, maintenance materials, and resins used for cooling water purification.

Occupational Safety and Health Administration

The Occupational Safety and Health Act (OSHA) covers a broad range of issues relating to worker health and safety, many of which impact the power generation industry (16,17). The Act sets standards designed to protect worker health and safety, particularly in industrial settings. The Occupational Health and Safety Administration, organized under the U.S. Department of Labor, implements and enforces OSHA standards and periodically updates policies governing worker health and safety.

The administration's rules cover many aspects of the power plant working environment, including specification of acceptable work garments and footwear, safety equipment and procedures, and work exposure to noise and substances which can negatively impact health. For example, the administration has stringent regulations regarding worker exposure to known carcinogens, such as asbestos, a naturally occurring, fibrous silicate material. This example is appropriate because asbestos-containing materials have been widely used in many existing power generation facilities, particularly for insulating materials surrounding piping, boilers, and other hot components. In addition, asbestos was widely used for fireproofing in or on building materials such as floor tiles, wallboards, and ceiling tiles. As power plants undertake maintenance and repair procedures, asbestos-containing materials encountered must be properly handled, removed, or encapsulated in place to comply with OSHA, EPA, and local rules.

In 1987, the administration tightened the permissible exposure limit (PEL) for asbestos from 2.0 fibers per cubic centimeter (f cm³) to 0.2 f cm³ based on an eight-hour, time-weighted average (TWA) exposure. In addition, more stringent notification, training, and supervision requirements have been established at that time, which are being further revised. In 1988, the administration amended the exposure standards to establish an excursion limit of 1.0 f cm³ as averaged over 30 minutes. In October of 1995, the TWA exposure standard was further reduced to 0.1 f cm³, but the excursion limit remained the same.

BIBLIOGRAPHY

"Regulatory Agencies" in *ECT* 3rd ed., Vol. 20, pp. 108–127, by N. R. Passow, C-E Lummus.

1. *Federal Regulatory Energy Commission 1993 and 1994 Annual Reports*, Federal Regulatory Energy Commission, Washington, D.C., Aug., 1994 and Aug. 1995, respectively.
2. L. S. Hyman, *America's Electric Utilities: Past, Present, and Future*, Public Utilities Reports Inc., 1983.
3. *EEl Pocketbook of Electric Utility Statistics*, Edison Electric Institute, Washington, D.C.
4. *Electric Power Annual 1993*, Energy Information Administration, U.S. Department of Energy, Washington, D.C., 1978.
5. *Federal Energy Regulatory Commission Fact Sheet*, Federal Energy Regulatory Commission, Washington, D.C., Mar. 29, 1995.
6. *Steam-Electric Plant Operation and Design Report*, Form E1A-767, Energy Information Administration, U.S. Department of Energy, Washington, D.C.
7. *Air Pollutant Emissions Factor*, AP-42 Release IV, Environmental Protection Agency, Washington, D.C., July 1993.
8. *Clean Air Act Amendments of 1990: Conference Report*, U.S. Government Printing Office, Washington, D.C., 1990.
9. *SO₂ and NO_x Emissions Control Technology: Comparative Analysis of Options*, SFA Pacific Inc., Mountain View, Calif., Sept. 1991, pp. 4.1–4.8.
10. *National Annual SO₂ Emissions Trends, 1995–2015*, EPA-454-R-95-001, Environmental Protection Agency, Washington, D.C., June 1995.
11. *New York State Energy Plan*, Vol. 2, Issues Report, New York State Department of Environmental Conservation, New York State Public Service Commission, and New York State Energy Office, Albany, N.Y., Oct. 1994, pp. 76–80.
12. *The SFA Quarterly Report*, SFA Quarterly Fuels and Feedstock Program, SFA Pacific Inc., Mountain View, Calif., Nov. 1994.
13. J. S. Walker, *A Short History of Nuclear Regulation, 1946–1990*, Nuclear Regulatory Commission, Rockville, Md., Jan. 1993.
14. *NRC Information Digest*, Nuclear Regulatory Commission, Rockville, Md., Jan. 1993.
15. *NRC Annual Report, 1992*, Nuclear Regulatory Commission, Rockville, Md., July 1993.
16. *Asbestos Control and Replacements Guidelines for Electric Utilities*, Electric Power Research Institute, Palo Alto, Calif., and Empire State Electric Energy Research Corp., N.Y., Aug. 1994.
17. E. Stern, personal communication, Occupational Safety and Health Administration; *Asbestos Advisor* (Expert System software), Release 1.0, OSHA, Washington, D.C., 1996 (available at <http://www.osha.gov>).

General References

"CAA Phase I, Special Report," *Elect. World*, 23–25 (Feb. 1994).

"Responding to the Clean Air Challenge," *EPRI J.* 21–29 (Apr. May 1991).

"Clean Air Act Amendments Update: Utility NO_x Regulations," EPRI NO_x Control Workshop, Scottsdale, Ariz., May 1994.

Steven Collins
Empire State Electric Energy Research Corporation

REINFORCED PLASTICS

Reinforced plastics are a relatively small, but very important part of the total plastics market. In 1994, U.S. production was approximately 1.4 billion kg, ~4% of total U.S. plastics shipments. Reinforced plastics are commonly referred to as composites or more specifically polymer composites, although it should be noted that not all composites are reinforced plastics. Ceramic metal-matrix composites and concrete are good examples of nonpolymeric composites (see COMPOSITE MATERIALS). A distinction is also drawn between reinforced plastics and advanced composites based on strength and performance. Reinforced plastics are also referred to as RP, FRP (fiber glass-reinforced plastic), and GRP (glass-reinforced plastic) interchangeably. The term fiber glass often is used to mean fiber glass-reinforced plastic, although it can also refer to insulation, textiles, and other forms of glass fibers.

Common to all reinforced plastics are two ingredients, resin and reinforcement. Resin is an organic material, usually of high molecular weight, that can be molded and set into a final shape. Resins are of two basic types. Thermoplastic resins soften upon heating, are shaped in a mold, and retain that shape when cooled. Common examples are nylon, polyethylene, polypropylene, and polycarbonate. Thermosetting resins are placed in a mold and cured by the use of a catalyst, heat, or both, until they harden in the shape of the mold. Common examples are polyester, vinyl ester, epoxies, phenolics, and polyurethanes.

The second main ingredient in reinforced plastic is the reinforcement, eg, fibers of glass, carbon, boron, mineral, cellulose, or polymers. Reinforcements can be configured in many ways, such as continuous or chopped strands, milled fibers, rovings, tows, mats, braids, and woven fabrics.

Reinforced plastics may also include fillers (qv), which are inexpensive materials such as calcium carbonate used to displace resin and reduce cost; curing agents (catalysts), promoters, inhibitors, and accelerators, which affect thermosetting resin cure; colorants; release agents (qv) to facilitate removal from the mold; and other additives which can impart a wide variety of properties to the finished part, such as fire resistance, electrical conductivity, static dissipation, and ultraviolet resistance.

It is important to note that reinforced plastics remain a combination of materials differing in form or composition on a macro scale. The main constituents (resin, reinforcement, and filler) retain their identities and do not dissolve or merge into each other; rather, they act in concert. These components can be physically identified and exhibit an interface between each other.

Manufacturing Method Selection Factors

There are many different manufacturing methods available for reinforced plastics. The selection process for a given application requires simultaneous consideration of the product design; the constituent materials; required physical, mechanical, and chemical properties; and production requirements. Assistance in selecting a manufacturing process is available from reinforced plastics manufacturers, material suppliers, and industry organizations such as the Composites Institute of the Society of the Plastics Industry and the Composite Fabricators Association. Selection factors that help to define the appropriate processes to consider for a given application may be categorized as product design factors, material factors, mechanical property factors, and production factors.

<i>Product design factors</i>	thermal properties
overall part size	electrical properties
shape complexity	weather resistance
critical dimensions	fire performance
weight limitations	<i>Mechanical property factors</i>
potential for parts integration	
assembly requirements	reinforcement type, amount, and orientation
secondary operations	loads to be applied
<i>Material factors</i>	static
	dynamic
polymer type, thermoplastic or thermosetting	impact
	assembly
reinforcement type, amount, and orientation	shipping
surface requirements	deflection limitations
chemical exposure	operating temperature range
<i>Production factors</i>	production rate buildup over time
total design life	
	production rate (highest monthly rate)
volatility of design, design modifications	shipping range

All of these factors contribute to the manufacturing cost of the composite product and thereby determine not only the most cost-effective composite process to use but also the competitive position of composites with the other materials of construction.

In most cases volume requirement is the primary consideration for selecting the most cost-effective process. However, in many cases a lower volume process may be used during initial phases of a program and then be replaced by a higher volume process when the product has gained field acceptance.

Manufacturing Methods

Hand Lay-Up and Spray-Up. In hand lay-up, fiber reinforcements in mat or woven form are placed on the mold surface and then saturated with a liquid polymer, typically a polyester resin, that has been chemically activated to polymerize (cure) without the addition of heat. Multiple plies of reinforcement and multiple cure steps allow very heavy wall thicknesses to be achieved.

In the spray-up process a reinforcement, usually glass fiber, is substituted for the mat and a special spray gun simultaneously chops the glass fiber and applies it with catalyzed resin to the mold surface. Hand rolling techniques then consolidate the fiber and resin to conform to the mold surface contours. The shorter chopped fibers allow for more intricate design details than do mats. Both processes rely heavily on the operators' skills for product quality. These two processes require the least capital investment and have the largest product size capability of all the processes. A single-surface mold produces a part with one controlled (usually the visible) surface.

Gel coats are typically used to provide a part with a finished surface directly from the mold. Various inserts, stiffeners, and mechanical attachments can be incorporated in the molding step, thereby further reducing secondary operations. Final edge trimming is accomplished with a variety of tools such as

routers, water jet cutters, or abrasive grinders with fixtures to define required dimensions. Typical products made by lay-up and spray-up include boat hulls, tub shower units, furniture, playground equipment, building facia, truck body parts, and corrosion-resistant tanks.

Vacuum Bag, Pressure Bag, and Autoclave Molding. These thermoset processes are variations of the hand lay-up process where pressure is applied to the reinforcement and resin during the cure step. With vacuum bag molding, a plastic film is placed over the resin-saturated reinforcement and carefully sealed around the perimeter of the part to form the bag. Air is then drawn from within the bag, allowing atmospheric pressure to be applied to the bag and the enclosed saturated reinforcement. This pressure permits higher reinforcement content (increasing strength) and reduces the labor needed to consolidate the resin and reinforcement. Additionally, the laminate is void-free, which makes this a common process technique for aircraft and other high performance components. A variation of this process provides for the addition of resin after the vacuum has been drawn on the bag. This technique has the advantage of reducing the emission of volatile components of the resin into the workplace.

Pressure bag molding substitutes air or liquid pressure for the atmospheric pressure, allowing higher pressures and higher fiber contents to be achieved. Autoclave molding uses a pressurized heating chamber to apply both heat and pressure to cure the product after it has been prepared using the vacuum bag method. Entire aircraft wing and tail sections can be fabricated from reinforced plastics using very large autoclaves and carefully controlled heat and pressure cure cycles.

A special pre-impregnated (prepreg) reinforcement system usually used with these processes combines the reinforcement, typically in a cloth or tape form, with a polymer that has been partially polymerized (B-staged). The prepreg is cut into the required shape and positioned onto the mold surface prior to application of pressure and heat. Prepregs offer precise resin reinforcement ratios and control of end product properties. Prepregs of different reinforcements, such as glass, aramid, carbon, boron, and high molecular weight polyethylene, are available for a broad range of demanding performances needed in aerospace and sport equipment applications.

Resin-Transfer Molding and Cold Press Molding. Resin-transfer molding (RTM) and cold press molding are thermoset processes utilizing a clamping frame or press and matching male and female molds. The molds fully contain the reinforcement and resin during the consolidation and curing process which is accomplished at a relatively low pressure (<345 kPa (50 psi)). In both processes the reinforcement, either as a precut mat or preformed shape (preform), is placed in the open mold. In RTM the mold is closed and a catalyzed resin is then pumped into the mold cavity to saturate the reinforcement. The perimeter of the mold has a sealing gasket to contain the resin until it cures. In cold press molding the reinforcement and catalyzed resin are placed into the open mold and the resin is distributed through the reinforcement as the mold is closed and pressurized. Proper resin distribution in both processes requires experience in locating the injection port or in developing a resin pour pattern that results in a voidfree part.

The preform is a form of reinforcement also utilized in the other liquid molding processes, ie, structural reaction injection molding (S-RIM) and wet system compression molding. The preform is made in a separate manufacturing process where reinforcing fibers are chopped into 2.5–5 cm (1–2 in.) lengths and deposited on a screen shaped to conform with the part to be molded. Air is drawn through the screen to hold the chopped fibers in place while a binder, usually a liquid emulsion, is applied and dried to hold the chopped strands in the desired shape for placement into the final process mold. The binder also must resist the flow of resin in the fabrication process and keep the fiber orientation until the laminating resin has cured. An innovation to the preform process is robot-controlled fiber placement that results in rapid, uniform preforms for the liquid molding processes.

A special attribute of these processes is the ability to pre-position reinforcement, inserts, and core materials for stiffening ribs. Gel coatings can be applied to the mold surface to eliminate post-mold finishing. Because both surfaces of the part are formed in a mold to close tolerances, accurate assemblies are possible, which is a requirement for many automotive or truck body applications.

All forms and compositions of reinforcements, ie, mats, woven roving, glass, carbon, and aramid, are commonly used with these processes. Special continuous glass strand mats with a thermoplastic binder allow preforms to be made using thermoforming techniques. These processes are used for truck and autobody components, medical equipment cabinets, transportation seating, and other parts needed in the intermediate volume range (1,000–10,000 parts yr).

Reaction Injection Molding. Two variations of conventional reaction injection molding (RIM) are used for reinforced plastic parts: reinforced RIM (R-RIM) and structural RIM (S-RIM). With R-RIM, very short reinforcing fibers are added to one of the resin components prior to being injected under high pressure through an impingement mixer into the closed mold. The mold is held closed during injection by a special clamping device that can be articulated to orient the mold for proper injection and to open the mold for part removal. The reinforcement reduces shrinkage and thermal expansion, and improves thermal performance of the base polymer, typically a thermoset polyurethane. Principal reinforcements include chopped and milled glass fibers and wollastonite, a fibrous filler. Applications for R-RIM include automotive facia, bumper covers, and other high volume components.

S-RIM utilizes the same polymers, molds, and process equipment. However, the reinforcement is in the form of a mat or specially shaped preform. The reinforcement is positioned in the mold prior to mold closure and resin injection. This allows the use of a greater percentage of reinforcement and results in higher strength in the part. Longer fibers also add to strength improvement. S-RIM allows production of larger and more structurally demanding parts than would normally be made by unreinforced RIM, such as automotive bumper beams, pickup truck beds, and large facia components.

Compression Molding. Compression molding processes dominate the higher production volume reinforced plastics applications. These processes use a hydraulically operated press to form the part in matching metal molds and to hold the densified molding material in the desired shape until the resin system has cured (if a thermoset) or cooled sufficiently (if a thermoplastic) to permit part removal. The presses typically are capable of molding pressures of 1.7–24 MPa (250–3500 psi) and have produced fishing boats as long as 4.9 m (16 ft), as well as one-piece tilting truck hoods. Modern presses have systems that control closing speed and pressure for proper flow of the material system in the mold. Following are the four variations of compression molding defined by the form and nature of the material system used.

Bulk Molding Compound. BMC is a combination of a thermosetting resin, most commonly polyester, with chopped reinforcement, fillers, pigments, and catalysts that form a dough-like, ready-to-mold material system. BMC can be extruded into logs or hand-formed into mold charges of the proper weight to produce net shape parts with a very high material efficiency. Molding pressures ranging from 2.4 to 13 MPa (350–2000 psi) and temperatures between 121 and 177°C (250–350°F) result in cure cycles from 30 seconds to several minutes depending on the thickness of the product.

BMC parts can be very complex with intricate detail and close dimensional tolerance, which can be used in assemblies with little or no post-molding machining steps. Molded-in inserts for screws, bearings, or other attachments are commonly used. Pigmented BMC eliminates additional surface finishing operations. Primary BMC applications include circuit breaker bases, power tool housings, computer and business equipment bases, pump housings, appliance components, automotive heater and air conditioner housings, and many other high volume complex industrial components.

Sheet Molding Compound. SMC is a ready-to-mold material system that combines the reinforcement, thermosetting resin, fillers, pigments, catalysts, and other additives in a continuous sheet that is formable into complex shapes in a single molding step with minimal scrap material. SMC is produced on a continuous forming machine where the liquid and dry ingredients, with the exception of the reinforcement, are premixed to form a paste of 40,000–100,000 mPa·s(cP) . The paste is metered by means of doctor blades onto two separate carrier films that are pulled continuously through the SMC machine. The lower carrier film moves under a chopper unit that chops and distributes the fibers uniformly over the resin paste. The upper and lower carrier films with paste and reinforcement are brought together, forming a sandwich that progresses through a compaction section whose mechanical action intimately mixes the paste and reinforcement prior to being rolled into convenient lengths or festooned into containers. The SMC is then stored under controlled temperature conditions.

Thickening agents added to the paste cause the SMC to increase in viscosity after forming, a step called maturation. A viscosity from 6,000 to 75,000 Pa·s (60,000–750,000 P) prior to molding is critical for SMC to flow well in the molding process and provide uniform mechanical properties throughout the molded part. More recent SMC technology, called low pressure molding compound (LPMC), has been introduced which uses a different resin thickening mechanism that provides uniform flow at much lower viscosity and therefore requires much lower molding pressures.

SMC with chopped reinforcement produces a part with essentially uniform mechanical properties in all directions. The mechanical properties are highly dependent on the flow of the SMC in the mold. When the flow is uniform, the properties are uniform. Both product and mold design must consider flow. Computer programs have been developed to predict the flow of SMC during the molding cycle to provide guidance in design of SMC products. By adding continuous reinforcements to the SMC process, directional strength properties can be increased if required for a specific application. Automotive bumper beams have used this type of system to increase beam strength while still retaining the flow characteristics of SMC with chopped reinforcement.

A variety of thermosetting resins are used in SMC. Polyesters represent the most volume and are available in systems that provide low shrinkage and low surface profile by means of special additives. Class A automotive surface requirements have resulted in the development of sophisticated systems that commercially produce auto body panels that can be taken directly from the mold and processed through standard automotive painting systems, without additional surface finishing. Vinyl ester and epoxy resins (qv) are also used in SMC for more structurally demanding applications.

Molding of SMC requires that the carrier film be stripped and a mold charge prepared. Special equipment is available to accomplish this step with minimum labor and good weight control of the mold charge. The carrier film is removed and rolled up for recycling (qv). Molding temperatures range from 121 to 177°C (250–350°F) for typical polyester systems. Molding pressures may vary from 3.5 to 17 MPa (500–2500 psi) depending on the size and complexity of the part. Low pressure SMC systems require from 0.7 to 2.4 MPa (100–350 psi) for the same part.

SMC materials have been used primarily in high volume applications in the automotive, appliance, electrical, and construction markets because of the tooling cost and high capital investment of the SMC and molding equipment. Automotive grill opening panels, bumper beams, exterior body components, engine valve covers, radiator surrounds, and wheels are in production. Appliance applications include dishwasher inner doors, air conditioner base pans and bulkheads, and business equipment housings. Bath tubs, showers, and door panels are principal construction applications. Steel assemblies of many individual parts can be replaced by a single SMC molding, as has been demonstrated with automotive grill opening panels and commercial air conditioner enclosures.

Wet System Compression Molding. Wet system compression molding was the first high volume method for manufacturing reinforced plastic parts, in such applications as the Chevrolet Corvette, industrial trays, tote boxes, luggage, refrigerator liners, and other commercial applications.

The process is similar to cold press molding with the exception that heated metal molds and higher molding pressures (1.7–6.9 MPa (250–1000 psi)) are used. The reinforcements used in the wet system are either a preform or in a mat form. The ability to achieve higher reinforcement weight loading and to preposition the reinforcement makes this material option viable for high volume structural applications. Low surface profile resin technology has been adapted to the wet molding systems for such applications as outboard motor shrouds, garden tractor shrouds, and snowmobile bodies, where both surface appearance and structural demand are required.

Reinforced Thermoplastic Sheet. This process uses precombined sheets of thermoplastic resin and glass fiber reinforcement, cut into blanks to fit the weight and size requirements of the part to be molded. The blanks, preheated to a specified temperature, are loaded into the metal mold and the material flows under molding pressure to fill the mold. The mold is kept closed under pressure until the temperature of the part has been reduced, the resin solidified, and demolding is possible. Cycle time, as with thermosetting resins, depends on the thickness of the part and the heat distortion temperature of the resin. Molding pressures are similar to SMC, 10–21 MPa (1500–3000 psi), depending on the size and complexity of the part.

Polypropylene sheet has been used most extensively; however, thermoplastic polyester, polycarbonate, and nylon versions are available (see ELASTOMERS, SYNTHETIC; POLYCARBONATES). Continuous strand glass fiber mat is the typical reinforcement. The limited number of sheet suppliers reduces potential for competitive pricing.

The process provides fast molding cycles, unlimited shelf life for the sheet, large part capability, and design flexibility. The process also allows for scrap materials to be recycled. Trim waste from the molding operation and defective parts can be ground up and recycled into the basic sheet process in controlled amounts. Some of this waste has also been used as input for injection molding.

As with SMC, applications are limited to high volume because of the capital investment in equipment and tooling. Thermoset compression molders require additional heating and material handling equipment to adapt their process to thermoplastic sheet fabrication. Applications include automotive bumper beams, load floors, radiator supports, battery trays, and package shelves. Chair shells, military containers, material handling pallets, trays, and concrete foaming pans are also produced.

Injection Molding. Injection molding is a thermoplastic or thermoset process that provides very high volume production capability. The basic process for injection molding reinforced plastics is similar to unreinforced plastics (see PLASTICS PROCESSING). The reasons for reinforcement differ depending on the specific polymer, but generally reinforcement increases certain mechanical properties, improves dimensional stability, and increases heat resistance, allowing the reinforced product to operate at higher temperature.

Injection molding of thermoplastics involves feeding a granulated or pelletized molding material into a screw-operated heating chamber, melting the molding material as the heated chamber fills, feeding the melt into the closed metal mold by means of the screw, allowing the melt to cool and solidify in the mold, opening the mold, and ejecting the finished part. Thermosetting reinforced plastics in BMC form are also injection molded. Different machine specifications are required for BMC in order to minimize the damage to the reinforcing fibers during feeding and injection, and the mold must be heated for thermoset cure rather than cooled as with thermoplastics.

Fast molding cycles, very low labor requirements, and complex part geometry are the primary advantages of the injection molding process. Reinforced thermoplastic polymers include nylon, PVC, acetal, polycarbonate, polyethylene, polypropylene, ABS, and SAN. Thermosets include polyester, epoxy, phenolic, and urethane. Reinforcements are typically chopped glass fibers from 3–13 mm (0.125–0.5 in.). Applications for injection molding include a wide range of automotive and appliance parts, electrical components, consumer products, and other markets that are served by unreinforced injection molding.

Pultrusion. Pultrusion is the reinforced plastic process that produces continuous profiles. In this process, reinforcements are pulled through a liquid thermoset resin bath into a heated mold where the cure reaction occurs. The pulling device has specially shaped gripping surfaces that continuously pull the cured profile out of the mold, thereby activating and controlling the speed of the entire process. A cut-off device trims the pultruded product to the required lengths. The pultrusion process is highly automated and once the process is started, very little direct labor is required. Long production runs provide the best economics. Tooling and capital equipment costs for pultrusion are generally less expensive than in the other high volume output processes. Pultruded products have a constant cross section that can be solid or hollow. Pultrusion requires a high degree of fiber orientation in the machine direction in order to activate the process, move the product through the curing mold, and provide the high tensile strength needed for applications such as oil-well sucker rods.

Many different thermosetting polymers are used in pultrusion, eg, polyester, vinyl ester, epoxy, and urethane. Reinforcements must be in a continuous form such as rovings, tows, mats, fabrics, and tapes. Glass fibers are the low cost, dominant composition, but aramid and carbon fibers are also used.

Pultrusion using thermoplastic polymers is a relatively new process variation that promises to grow rapidly in the future. Higher processing speed, a wide variety of available thermoplastics and polymers, and potential for post-forming of the pultruded part has stimulated development of this process. Pultrusion products include fishing rods, ladder rails, electrical strain rods, concrete rebar, structural shapes (angles, I-beams, channels, etc), tubing, electrical ducting, cable tension members, archery arrows, and window and door lineals.

Filament Winding. This is a process for products that are surfaces of revolution. Reinforcing fibers are drawn through a liquid resin bath and applied to a rotating mold surface or mandrel. For many applications, the reinforcement is precombined with the liquid resin to form a prepreg product. Prepregs are used for the most critically demanding applications, such as rocket motor cases and pressure tanks, because of the higher degree of product and process control they provide. By varying the rotating speed of the mandrel and the movement of the fibers across the mandrel, geometric patterns are formed that orient the fibers to meet the product strength requirements. The principal thermoset polymers used in this process are polyesters, vinyl esters, bisphenol fumarates, epoxies, furans, and phenolics. More recently, thermoplastics are also used in the form of roving or tape prepregs. These materials require localized heating during winding to consolidate the tape layers. Reinforcements for this process include continuous strands and tows, woven and unidirectional tapes, and chopped and continuous strand mats. Glass fiber is the dominant reinforcement with carbon, aramid, and high molecular weight polyethylene also used for the structural or weight-critical applications.

Sophisticated structural analysis techniques make it possible to determine both the amount and exact orientation of reinforcement that the product will need to meet the critical stresses in actual service. Hybrid reinforcement systems containing different fiber compositions with different properties are being increasingly used. For example, hybrid carbon and glass fiber automotive drive shafts are in commercial use.

There are many different equipment options available to suit specific product needs including continuous winders for pipe, multiaxis winders for pressure vessels, and simple lathe-type winders for tanks and large pipe. Specialty machines combine a chopped reinforcement with continuous fibers for tank walls and large-diameter pipe where both stiffness and tensile strength are required. Textile braiders have also been adapted for use as continuous

winders where both longitudinal and helical winding patterns can be formed. Although limited to surfaces of revolution, filament winding represents one of the most efficient reinforced plastic processes because of the highly controllable properties, and ability to achieve high strength performance at the lowest possible weight.

Products can be found in every principal market area including rocket motor and shell casings, air and gas pressure tanks, aircraft wing fuel tanks, utility poles, automotive and truck drive shafts, sailboat masts, vaulting poles, fishing rods, golf shafts, railroad tanks cars, and pipes and tanks for oil, gas, and chemical processing.

Other Processes. There are a variety of other processes that have been developed to suit specific product needs.

Continuous laminating produces flat and corrugated sheet used as translucent glazing, signs, wall covering, and recreational vehicle siding. Glass fibers are chopped onto a moving film and then saturated with resin. Another film is applied to the surface before shaping and curing in an oven.

Centrifugal casting is used to produce water softener tanks and pipe by saturating a reinforcement with thermosetting resin within a mold that is then rotated at high speed to consolidate the laminate before curing.

Rotational molding is used to form large shells of thermoplastic resin and chopped strands for such applications as agricultural tanks and fertilizer hoppers. The resin and chopped glass are placed in the metal mold that is then rotated in an oven where the thermoplastic resin melts and deposits the fiber on the metal surface. When cooled, the mold is opened and the part is removed.

The *rigidized thermoplastic sheet* process combines thermoplastic sheet vacuum forming with a spray-up or cold press molding process to add a thermoset composite structural backing to a decorative thermoplastic skin. Large parts such as bathtubs, hot tubs, recreation vehicle components, and camper tops have been produced by this process.

Hybrid Processes. There are also hybrid processes that have evolved to meet specific product needs. As an example, automotive leaf springs utilize a filament winding system to prepare impregnated fiber bundles that are then compression molded to final configuration.

Economic Aspects

The business climate of the 1990s is different from the past. Factors such as increased competition, a global marketplace, rapid technical shifts, and greatly compressed product life cycles constantly open new opportunities for plastics in general and reinforced plastics in particular. Reinforced plastics have become widely accepted for particular applications because they offer a combination of design, performance, and economic benefits to the user. These materials have had a proven record of success since the 1940s.

Totally new product applications, such as in aerospace, for example, are a minority of the successful applications that constitute the composite industry, and these uses tend to be small niche markets that are cost-insensitive. In most situations composite materials are substituted for existing materials, most often wood and common metals. Traditional materials, however, are often entrenched and hard to displace. There are established product specifications, widely accepted traditional economics, and often a long history of industry experience. The result is that the product requirements are usually a material rather than performance specification, and moreover they are usually established in terms of existing materials, thereby providing a barrier for substitution. Furthermore, if just a component of a larger system is being considered for substitution, other limitations on shape, size, choice of materials, etc, probably exist that prevent maximum benefit of composites from being realized. Nevertheless, in spite of these concerns, composites have been successful since the 1940s because of the benefits offered over the existing material. It is estimated that there are over 40,000 different applications for reinforced plastics.

Design Benefits. Design benefits include ways of making a finished part that eliminate fabrication steps needed for traditional materials. Parts with complex surface shapes are often good candidates for composite materials because composites can be made to virtually any predetermined shape from a mold. No cutting, stamping, shaping, or further processing is needed. Parts consolidation is another benefit that favors reinforced plastics. Composites can be designed in complex arrangements that can replace assemblies of many metal parts and fasteners, greatly reducing assembly labor costs and resulting in a better quality part. An example is in automotive front-end assemblies, where a single reinforced plastic piece has replaced a dozen or more metal pieces and fasteners.

Composites are superior materials for applications that require large part size. Large boat hulls, for example, can be molded in a single piece, eliminating costly cutting and fastening operations needed with wood or metal. Composites generally entail reduced finishing costs. They are easier to design into a finished product than a bulk material, such as lumber or rolled steel, which has to be cut, shaped, fit together, fastened, sanded, painted, and otherwise finished.

Strength can be engineered to the magnitude and in the direction needed for the application. Strength can be tensile, such as oil-field sucker rods; flexural, such as sporting goods; compressive, such as large pipes; or impact, such as in automotive wheels. Stiffness can also be engineered as needed. Composites generally have high impact strength that allows them to absorb a large amount of energy before failure. They also lack ductility so that products, when properly designed, maintain alignment through their life cycle.

Some design factors, however, work against composites. For example, glass fiber-reinforced plastics generally have lower modulus (stiffness) than metals. Thickness and shape adjustments are required where stiffness is a critical design requirement. With appropriate reinforcement, any modulus, even greater than that of metals, can be achieved. However, it may become expensive and uneconomical to do so.

Service temperature limitations must be considered in the use of composites, not only in the selection of polymer and process, but sometimes in the selection of the reinforcement as well. Composites cannot generally perform as well as metals or ceramics in very high temperature applications, but they can be made fire resistant to meet most construction and transportation codes.

Performance Benefits. Fiber-reinforced plastic (FRP) components are lightweight and deliver more strength per unit of weight, or less weight for a given strength requirement than unreinforced plastics and most metals. In transportation applications light weight means better fuel economy, and in marine and aerospace usage, light weight is critical for flotation and aerodynamic performance, respectively.

Reinforced plastics provide long-term corrosion resistance to many chemical and temperature environments, making them especially suited for chemical and power plants, and oil-field collection and distribution. Composites can have high dielectric strength. They are usually excellent electrical insulators and do not absorb moisture, making them suitable for use in such products as ladders and printed circuit boards because they provide a stable, nonconducting structure. Composites can have low thermal conductivity, making them suitable in applications that call for thermal insulation. An FRP bathtub, for example, retains heat longer than a cast-iron tub.

Under mechanical and environmental stresses, composites are dimensionally stable. They maintain their shape and functionality, a critical requirement in such applications as dish antennas, construction girders, and in appliance and business machines. Color and surface texture can often be molded into an FRP product for long lasting, low maintenance permanent surface appearance. Boats are a good example. The surface color is molded in and requires minimum maintenance, an advantage in saltwater environments.

Economic Benefits. The traditional costs of a product include raw materials, processing, overhead, and so forth. Designers and engineers considering potential composite applications cannot compare material costs only. Polymer composites, except for inexpensive fillers and small amounts of additives, consist mostly of resin and reinforcement, whose materials costs are usually higher than traditional materials.

Process costs must also be carefully considered. Almost all the composite processes are batch processes, producing one or at most a few parts at a time. Pultrusion and some pipe winding processes are exceptions; even though these processes are continuous, they are relatively slow as compared to plastic extrusion or some of the metal fabrication processes. Molding times for composite processes range from less than one minute to hours. Although productivity may appear to put composites at a disadvantage compared to faster manufacturing methods, such as metal stamping, the design benefits usually more than compensate in reduced assembly and finishing work required. Moreover, tooling costs can be lower for composite molding than for metalworking, especially when they need to be allocated over a limited number of parts.

Designers and engineers have successfully substituted composites into many applications because not only traditional costs, but the overall life cycle economics of the product in the manufacturing process and in the product's later application have been considered. Designers and engineers recognize that the real cost of a product is the total cost of everything associated with it. Life cycle economics thus include nontraditional costs such as planning, design,

and development costs; purchasing, installation, and maintenance costs; loss and wear costs; liability and insurance costs; downtime and lost business costs; and replacement, disposal, and recycling costs.

The difference between life cycle costs and the traditional accounting cost is that traditional costs are fairly easy to quantify, whereas life cycle costs are usually difficult to quantify and are usually either lumped into general overhead costs or ignored altogether. They are very real, however, and when examined in detail and fully quantified, the economics of composites allow them to displace the traditional materials.

There are no simple rules of thumb in defining the cost of reinforced plastic components. Their successful use has resulted from proper design, utilizing the benefits these materials offer, process selection, tooling cost advantages that fit the production needs, and consideration of life cycle economics. Each existing application illustrates the cost-performance advantage of reinforced plastic over the traditional material that is displaced.

BIBLIOGRAPHY

"Lamination and Laminated Products" in *ECT* 1st ed., Vol. 8, pp. 185–192, by H. W. Narigan and G. E. Vybiral, St. Regis Paper Co.; "Laminated and Reinforced Plastics" in *ECT* 2nd ed., Vol. 12, pp. 188–197, by C. S. Grove, Jr., Syracuse University and D. V. Rosato, Plastics World; in *ECT* 3rd ed., Vol. 13, pp. 968–978, by F. J. McGarry, Massachusetts Institute of Technology.

General References

Books

K. H. G. Ashbee, *Fundamental Principles of Fiber Reinforced Composites*, Technomic Publishing Co., Inc., Lancaster, Pa., 1989.

M. L. Berins, ed., *Plastics Engineering Handbook*, 5th ed., Society of the Plastics Industry, Inc., Van Nostrand Reinhold, New York, 1991.

Composites: An Insider's Technical Guide to Corporate America's Activities, Turner Moss Co., New York, 1989.

Facts and Figures of the U.S. Plastics Industry, 1995 ed., Society of the Plastics Industry, Inc., Washington, D.C., 1995.

Introduction to Composites, 3rd ed., Society of the Plastics Industry, Inc., Washington, D.C., 1995.

G. Lubin, ed. *Handbook of Composites*, Van Nostrand Reinhold, New York, 1982.

R. H. Sonneborn, A. G. H. Dietz, and A. S. Heyser, *Fiberglass Reinforced Plastics*, Reinhold Publishing Corp., New York, 1954.

Trade Journals

Composites Design & Application, Composites Institute of the Society of the Plastics Industry, Inc., New York, published monthly.

Composites Fabrication, Composites Fabricators Association, Arlington, Va., published monthly.

Composites News: Infrastructure, Composites Worldwide, Inc., Solana Beach, Calif., published biweekly.

Composites Technology, Ray Publishing, Inc., Denver, Colo., published bimonthly.

High-Performance Composites, Ray Publishing, Inc., Denver, Colo., published bimonthly.

Modern Plastics, McGraw-Hill, Inc., New York, published monthly.

PlasticsBRIEF, Reinforced Plastics Newsletter, Market Search, Inc., Toledo, Ohio, published biweekly.

Plastics Distributor & Fabricator, PMD Published Inc., Chicago, Ill., published bimonthly.

Plastics Engineering, Society of Plastics Engineers, Inc., Brookfield, Conn., published monthly.

Plastics News, Crain Communications, Akron, Ohio, published weekly.

Plastics Technology, Bill Communications, Inc., New York, published monthly.

Plastics World, PTN Publishing Co., Melville, N.Y., published monthly.

Reinforced Plastics, Elsevier Advanced Technology, Oxford, U.K., published monthly.

Andrew L. Bastone
Steven G. Katz
ISORCA, Inc.

RELEASE AGENTS

Release agents are substances that control or eliminate the adhesion between two surfaces. They are used to expedite many industrial handling and processing operations, particularly of polymers, such as calendering, casting, embossing, extrusion, forming, labeling, laminating, machining, molding, packaging (qv), protecting, and transferring (see PLASTICS PROCESSING). They are known by a variety of terms descriptive of their effect, including abherents, abhesives, antiblocking agents, antistick agents, external or surface lubricants, mold-release agents, parting agents, and slip agents. They find considerable use in the adhesive, food, furniture, glass, metal, plastics, and rubber industries.

Release agents function by either lessening intermolecular interactions between the two surfaces in contact or preventing such close contact. Thus, they can be low surface-tension materials based on aliphatic hydrocarbon, fluorocarbon groups, or particulate solids. The principal categories of material used are waxes, fatty acid metal soaps, other long-chain alkyl derivatives, polymers, and fluorinated compounds.

Some of these processing aids are incorporated into the bulk of the material rather than being directly applied to the surface and are known as internal release agents. In the process of migration to the surface they have an inevitable internal lubrication effect. Although it is a common practice to classify lubricant materials such as the metallic soaps as internal or external, most internal lubricants also have an external lubricant effect, particularly at higher concentrations because usually some material finds its way to the exterior. As of 1996, consumption of externally applied release agents far outweighs the use of internal release agents.

Product Types and Requirements

Release agents are available in a wide variety of forms and are formulated for numerous modes of application. Product types include neat liquids, solutions, powders, flakes, pastes, emulsions, dispersions, sprays, and films. Some are general-purpose inert products intended for a broad range of applications, including home consumer uses, whereas others are highly specific, designed to react *in situ* with particular substrates. As with most other formulated processing aids, there is a trend away from products containing volatile organic compounds (VOCs) and ozone-depleting substances, and toward water-based and high solids formulations. Users are also switching from the heavier metal soaps, such as lead stearate [7428-48-0]. Another trend that helps reduce solvent and carrier use is toward semipermanent release treatments for multiple releases that have to be applied less frequently. Many products serve more than one processing function and contain other additives, such as antioxidants and wetting agents. These multifunctional and reactive products are usually proprietary with little information available about their composition.

Suppliers of release agents vary from large, basic polymer producers to small, regional formulators and distributors. Examples from among the hundreds of such suppliers include the following:

- Acmos Inc.
- Air Products and Chemical
- American Cyanamid Chemical Products Division
- Bareco Division, Petrolite
- Chem-Trend, Inc.
- The Dexter Corporation
- Dow Corning Corporation
- E. I. du Pont de Nemours & Co. Inc.
- Emery Industries
- Th. Goldschmidt AG
- C. P. Hall Company
- Kluber Chemie KG
- Mallinckrodt Specialty Chemical Company
- Rhône-Poulenc
- Witco Corporation

Many more suppliers and information about their product line can be obtained from compilations such as the *Modern Plastics Encyclopedia* (1) and the *Thomas Register of American Manufacturers* (2). The choice of a release agent depends on the process conditions involved and the nature of the contacting substrates (3). Apart from the obvious ease of release, other important requirements are minimal buildup of residues on mold substrate, minimal effect on the molded article, adequate film-forming ability, compatibility with secondary operations and other processing parameters, health and safety requirements, and cost.

According to a survey of polymer processors in the United States (4), the prevention of residue buildup is the most common shortcoming of external release agents. This residue can occur from either the mold release or the resin or molding compound. Should buildup occur, a secondary requirement is ease of cleanability of the mold substrate. Undesirable effects on the properties of the surface of the molded article include blistering, streaking, discoloration, stress cracking, and solubility in the plastic. The ability to form and maintain an adequate film is critical in operations such as casting and molding, where movement of the melt might displace the release agent. Provision of sufficient cohesive strength of the film and substantivity to the mold can be achieved by incorporation of polar groups such as amide or carboxylate, but the conflicting need to present no adhesion-promoting groups at the interface with the melt must also be considered. One key processing parameter is the time available for the release to occur. For example, the carrier, whether solvent or water, must completely evaporate before a mold is filled. This can be critical in automated processes and limits choices of release agents. Similarly, many applications are at elevated temperatures, and good thermal and oxidative stability are important to avoid formation of undesirable, polar, adhesive-like groups. Health and safety requirements are specific to the use; the most stringent requirements are in the preparation, processing, packaging, and applications of foodstuffs and drugs. Cost effectiveness or number of suitable releases per unit cost are the real cost requirements to be considered.

These criteria guide product selection for a particular application, but trials under production conditions are usually necessary. Some standard test methods are available that closely simulate actual end-use operations. For example, ASTM D3354-89 (5) quantifies the degree of blocking, ie, unwanted adhesion, existing between layers of plastic film by measuring the load required to separate objects attached to parallel aluminum plates. There are also standard test methods (D1894-90 and D3028-90) for determining coefficients of friction of plastic solids. Such data can be very sensitive to film age and surface condition, but correlation with actual performance can usually be established. Surface appearance is inevitably affected by the presence of these additives, and there are standard tests for surface gloss and haze, eg, D2457-90 and D1003-92.

The effects of release additives on bulk properties must also be carefully considered, particularly with integral additives to plastics. For example, partial solubility usually confers some plasticizing effect. This may improve impact strength but could reduce the heat distortion temperature. Some release additives such as metallic soaps have secondary antioxidant and heat-stabilizer benefits. Such effects are exploited in multipurpose formulations.

Classification of Release Agents

The diversity of release products and the wide range of release problems make classification difficult. One approach is by product form, with subdivisions such as emulsions, films, powders, reactive or inert sprays, reactive coatings, and so on. Another approach is by application, eg, metal casting, rubber processing, thermoplastic injection molding, and food preparation and packaging.

A classification by chemical type is given in Table 1. It does not attempt to be either rigorous or complete. Clearly, some materials could appear in more than one of these classifications, eg, polyethylene waxes [9002-88-4] can be classified in both synthetic waxes and polyolefins, and fluorosilicones in silicones and fluoropolymers. The broad classes of release materials available are given in the chemical class column, the principal types in the chemical subdivision column, and one or two important selections in the specific examples column. Many commercial products are difficult to place in any classification scheme. Some are of proprietary composition and many are mixtures. For example, metallic soaps are often used in combination with hydrocarbon waxes to produce finely dispersed suspensions. Many products also contain formulating aids such as solvents, emulsifiers, and biocides.

Table 1. Chemical Classification of Release Agents

Chemical class	Chemical subdivisions	Specific examples
waxes	petroleum waxes	paraffin wax [8002-74-2], microcrystalline wax [6474-42-3]
	vegetable waxes	carnauba wax [8015-86-9]
	animal waxes	lanolin [8006-54-0]
	synthetic waxes	polyethylene wax [8002-88-4], Fischer-Tropsch wax [68649-50-3]
fatty acid metal soaps	metal stearates	magnesium stearate [557-04-0], zinc stearate [557-05-1]
	other	calcium ricinoleate [6865-33-4]
other long-chain alkyl derivatives	fatty esters	diethylene glycol monostearate [106-11-6], hydrogenated castor oil [8001-78-3]
	fatty amides and amines	ethylenebis(stearamide) [110-30-5], oleyl palmitamide [16260-09-6]
	fatty acids and alcohols	palmitic acid [57-10-3], oleic acid [112-80-1]
polymers	polyolefins	polypropylene [9003-07-0]
	silicones	polydimethylsiloxane [9016-00-6], polymethyl(nonafluorohexyl)-siloxane [115287-18-8]
	fluoropolymers	polytetrafluoroethylene [9002-84-0], poly(fluoroethers)
	natural polymers	cellophane [9005-81-6]
fluorinated compounds	other	polyoxalkylenes, poly(vinyl alcohol) [9002-89-5]
	fluorinated fatty acids	perfluorolauric acid [307-55-1]
inorganic materials	silicates	talc [14807-96-6]
	clays	kaolin [1332-58-7], mica [12001-26-2]
	other	silica [7631-86-9], graphite [7782-42-5]

Mechanism of Release

Because release agents are used to reduce the adhesion between materials, knowledge of the mechanism of adhesion and the properties that affect it is essential to an understanding of the release process. The main mechanisms of adhesion are interdiffusion, electrostatic attraction, surface energetics and wettability, and mechanical interlocking. The important surface and interfacial properties include surface topography (composition, structure, and roughness), surface tension or energy and its effect on substrate and adhesive wettability, the thermodynamic work of adhesion, and chemical reactivity at the interface. Nonsurface properties of the thin-film interfacial phase include its ability to set to a cohesive solid after wetting the substrate and its viscoelastic response to deformation controlled by factors such as degree of crystallinity, molecular weight and distribution, number of cross-links, and presence of fillers. Surface treatments that enhance adhesion do so by removing weak boundary layers, changing surface topography, or changing the chemical nature of the surface and by introducing strongly attracting functional groups.

An inversion of these arguments indicates that release agents should exhibit several of the following features: (1) act as a barrier to mechanical interlocking; (2) prevent interdiffusion; (3) exhibit poor adsorption and lack of reaction with at least one material at the interface; (4) have low surface tension, resulting in poor wettability, ie, negative spreading coefficient, of the release substrate by the adhesive; (5) low thermodynamic work of adhesion; (6) low intermolecular forces across the interface, eg, an absence of electrostatic or polar attractions; (7) display nonsetting or low cohesive interactions within the release phase; and (8) provide a weak boundary layer.

Many of these features are interrelated. Finely divided solids such as talc [14807-96-6] are excellent barriers to mechanical interlocking and interdiffusion. They also reduce the area of contact over which short-range intermolecular forces can interact. Because compatibility of different polymers is the exception rather than the rule, preformed sheets of a different polymer usually prevent interdiffusion and are an effective way of controlling adhesion, provided no new strong interfacial interactions are thereby introduced. Surface tension and thermodynamic work of adhesion are interrelated, as shown in equations 1, 2, and 3, and are a direct consequence of the intermolecular forces that also control adsorption and chemical reactivity.

The work of adhesion, W_A , is the change in energy per unit surface area when two interfaces come into contact, as given in equation 1 where σ_1 and σ_2 are the surface energy of each phase and σ_{12} the interfacial energy between them.

$$W_A = \sigma_1 + \sigma_2 - \sigma_{12}$$

(1)

For liquid systems these surface energies expressed in mJ m² are numerically equivalent to the surface tensions in mN m (= dyn cm). If the adhesive is phase 1 and the release coating is phase 2, then the spreading coefficient, S , of 1 on 2 is as given in equation 2.

$$S = \sigma_2 - \sigma_1 - \sigma_{12}$$

(2)

Because the work of cohesion, W_C , of the adhesive is 2 σ_1 , equation 3 follows.

$$W_A = W_C - S$$

(3)

This simple yet fundamental relationship, first noted by Harkins (6), states that when the forces of adhesion are less than those of cohesion, ie, when W_A is less than W_C , the spreading coefficient is negative and failure will be at the interface between the adhesive and release coating, the desired situation in a release coating application. Conversely, when the spreading coefficient is positive, undesirable cohesive failure in the adhesive or release coating will be the case. These equations clearly demonstrate the importance of the release coating surface tension being lower than that of the adhesive.

The intermolecular forces of adhesion and cohesion can be loosely classified into three categories (7): quantum mechanical forces, pure electrostatic

forces, and polarization forces. Quantum mechanical forces give rise both to covalent bonding and to the exchange interactions that balance the attractive forces when matter is compressed to the point where outer electron orbits interpenetrate. Pure electrostatic interactions include Coulomb forces between charged ions, permanent dipoles, and quadrupoles. Polarization forces arise from the dipole moments induced in atoms and molecules by the electric fields of nearby charges and other permanent and induced dipoles.

The forces involved in the interaction at a good release interface must be as weak as possible. They cannot be the strong primary bonds associated with ionic, covalent, and metallic bonding, which have energies in the range of 80–200 kJ mol (19–48 kcal mol); neither are they the stronger of the electrostatic and polarization forces that contribute to secondary van der Waals interactions. Rather, they are the weakest of these types of forces, the so-called London or dispersion forces that arise from interactions of temporary dipoles caused by fluctuations in electron density. They are common to all matter and range from 0.1 to 40 kJ mol (0.02–9.56 kcal mol). The surfaces that are solid at room temperature and have the lowest dispersion-force interactions are those comprised of aliphatic hydrocarbons and fluorocarbons. This is reflected in critical surface tension of wetting values (8) and other ways of quantifying solid-surface energy where the lowest energy surfaces are those based on such constituent groups as CF₃, CF₂, CH₃, and CH₂ and the various mixed functions, CFH, etc.

Among the hydrocarbons, the lowest surface tension values are found for surfaces comprising closely packed methyl groups. This is a characteristic of waxes and oriented long-chain alkyl derivatives, accounting for their predominance among organic release agents. Because of the extreme localization of the fields of force of such covalently bonded atoms, the pendent methyl groups on polydimethylsiloxane [9016-00-6] also behave as an array of close-packed methyl groups, with little direct effect from the siloxane backbone. Aliphatic fluorocarbons are the lowest surface-energy solids known. Table 2 lists the surface tensions of some common release materials.

Table 2. Surface Tensions of Various Release Substrates

Substrate	CAS Registry Number	Surface tension, ^a mN m(dyn cm)	Reference
carnauba wax		38 ^b	9
paraffin wax		23 ^c	10
octadecylamine monolayer	[124-30-1]	22	11
stearic acid monolayer	[57-11-4]	21	11
polyethylene	[9002-88-4]	31 ^c	12
polypropylene		29 ^c	13
polydimethylsiloxane film		24 ^c	14
polydimethylsiloxane fluid		20.4 ^d	15
polymethyl(3,3,4,4,5,5,6,6-nonafluorohexyl)siloxane		16.3	16
poly(vinyl fluoride)	[24981-14-4]	28 ^c	12
poly(vinylidene fluoride)	[24937-79-9]	25 ^c	12
polytetrafluoroethylene		18.5	12
10-monohydroperfluoroun-decanoic acid monolayer	[1765-48-6]	15 ^c	14
poly(1,1-dihydroperfluoro-octyl methacrylate	[29014-57-1]	10.6	12
perfluorolauric acid monolayer		6 ^c	14

^a Critical surface tension of wetting, *n*-alkane test liquids only, unless otherwise indicated.

^b Extrapolated value from higher temperature measurements on liquid material.

^c Critical surface tension of wetting, mixed series of test liquids.

^d Direct liquid surface tension measurement.

Because the forces of attraction prevail when molecules are brought into sufficiently close proximity under normal conditions, release is best effected if both the strength of the interaction and the degree of contact are minimized. Aliphatic hydrocarbons and fluorocarbons achieve the former effect, finely divided solids the latter. Materials such as microcrystalline wax [64742-42-3] and hydrophobic silica [7631-86-9] combine both effects. Some authors refer to this combined effect as the ball bearing mechanism. A perfluoroalkylated fullerene nanosphere would perhaps be the ultimate example of this combined effect (17). These very general mechanistic remarks can be supplemented by publications on the mechanism of specific classes of release agents such as metallic stearates (18), fatty acids and fluorinated compounds (19), and silicone-coated release papers (20,21). The mechanism of release of certain problem adherents, eg, polyurethanes, has also been addressed (22,23).

Industrial Applications

Metal and Glass. Although it might be expected that only highly thermally stable release agents would be found to be in use with molten metals and glasses, this is not necessarily the case. For example, mineral oils are still widely used in molding and shaping glass, despite the considerable quantity of smoke and buildup of residues entailed. Heat-resistant, graphite-containing resins are also used. Phenolic resins are used in this application and also in mold fabrication for metal castings.

In this process, sand and phenolic and other resins are mulled together and applied to a pattern and hardened to give a mold. Release agents are needed to prevent adhesion of the resin binder to the pattern, thus enabling the mold to be cleanly slid away, and similarly to improve the mold release properties after the casting of the metal. Internal mold release agents such as calcium stearate [1592-23-0] have other beneficial effects, such as reducing agglomeration of the resin-coated particles and thereby providing a freer flowing molding sand as well as reducing muller power requirements. External release agents such as aqueous silicone emulsions or wax dispersions are used both in the mold-making and in the casting steps, depending on the temperatures involved. Similar release agents are used in a variety of other metal-forming operations, including the injection molding of low melting aluminum alloys, drawing of ferrous metals, and hobbing and stamping with application to both the tool and product surfaces.

Rubber and Plastics. Release agents are widely used in the rubber and plastic industry to achieve release of polymers and release from polymers. They are useful in many polymer-processing applications, such as extrusion, calendering, molding, and embossing, to prevent the polymer from sticking and building up on process equipment and thereby eliminating the accumulation of rejects. In subsequent machining, packaging, and labeling operations, abherents allow faster, more continuous handling. Release from polymer applications varies from the familiar examples of fluorocarbon-treated cookware to the use of shrink films of polytetrafluoroethylene [9002-84-0] or polyethylene [9002-88-4] over rollers used in the ink, printing, polymer processing, and coating industries. A significant aspect of this type of application is antideposition coatings to reduce accumulation of pollutants and other undesirable matter. Examples include nonthrombogenic surfaces in biomedical areas, and low soiling materials for solarenergy collection devices.

In the extrusion of thermoplastic polymer powder or granules, release agents aid in the flow and keep the finished product from sticking when stacked. They can also reduce the possibility of thermal breakdown by reducing the amount of energy required and providing shorter cycle times. Production of granules of particularly adhesive polymers such as polystyrene [9003-53-6] would be impossible without such aids. In the mixing, sheeting, and calendering of natural rubber [9006-04-6] sheets, release agents prevent stock from sticking to the rollers, thereby maintaining a continuous operation. Waxes, metal stearates, and silicones are widely used; solid abherents are popular for dusting finished, unfilled products.

Molding operations such as the manufacture of automobile seat foams and reactive injection molding (RIM) consume considerable quantities of release agents. In RIM, reactive liquid components are combined in a mix-head and injected into a mold to fabricate plastic parts. Many polymers are molded, but polyurethanes present one of the most significant challenges in this area, owing to the likelihood of adhesive bonding of isocyanates to the mold (22). Polymer films are often used to allow castings to separate from the mold. A specific example is the use of regenerated cellulose (cellophane [9005-81-6]) films in the manufacture of corrugated, glass-reinforced polyester building panels. The molds can also be made of polymers with appropriate release properties such as silicones. Additional advantages include easy processing, low shrinkage, good thermal resistance, good reproduction of detail, and light weight (24).

Adhesive Transfer Processes. Many polymers, whether deliberately or accidentally, are adhesives, so that much of the adhesive industry can be regarded as a part of the rubber and plastics industry. However, there are several important material-transfer applications involving polymer products that are so critically dependent on controlled adhesion that they merit specific mention in that category. They include hot stamping foils, release coatings for pressure-sensitive adhesive products, photocopier materials, transfer coatings, and transfer printing of textiles.

Hot stamping foils are a type of composite decorative laminate that uses a polymer carrier film, either polyester or cellophane, in conjunction with an external release agent, usually a wax, to transfer coatings to surfaces to be decorated. The transferred coating is rapidly bonded using heat and pressure. Particular attention must be paid to foil detail and surface area, nature of the transferred coating, and avoidance of premature release during slitting and winding of the foils. Some foils are metallized by vacuum deposition, hence there must be no volatiles or migration to interfere with subsequent adhesion on hot stamping.

Silicone products dominate the pressure-sensitive adhesive release paper market, but other materials such as Quilon (E.I. du Pont de Nemours & Co., Inc.), a Werner-type chromium complex, stearato chromic chloride [12768-56-8], are also used. Various base papers are used, including polyethylene-coated kraft as well as polymer substrates such as polyethylene or polyester film. Silicone coatings that cross-link to form a film and also bond to the cellulose are used in various forms, such as solvent and solventless dispersions and emulsions. Technical requirements for the coated papers include good release, no contamination of the adhesive being protected, no blocking in rolls, good solvent holdout with respect to adhesives applied from solvent, and good thermal and dimensional stability (see SILICON COMPOUNDS, SILICONES).

Transfer of toner material between various rollers and eventually onto the paper is a demanding process. Because speed is of the essence, a combination of temperature and pressure is used to ensure adequate flow of the thermoplastic toner when fusing the image. Temperatures close to 200°C are used, requiring good thermal stability of both the rollers and the release agents. To prevent melted toner from sticking to the heated roller, it is often coated with a low surface-energy material such as polytetrafluoroethylene, or silicone or fluorocarbon elastomers. Often a release agent or fuser oil is applied to the fuser roll to ensure adequate release. Silicone is the agent usually chosen, because of its excellent thermal and chemical stability. Another approach is to incorporate the release agent in the toner. Polyethylene and polypropylene [9003-07-0] waxes are dispersed in the toner and migrate to the surface during the fusing step.

Transfer coating is a combination of casting and coating used to create composite materials such as simulated suede leather. A polyurethane film is cast on a release sheet and dried; a second solution is cast on the dried film and then laminated to a suitable fabric while still wet. The release paper is then stripped away when the material has dried. Flexographic printing directly onto fabrics is beset by shrinkage and register problems that are considerably eased by transfer printing from a release sheet of abherent-treated paper. This is printed on the release side by the direct gravure method, dried, and rewound. Direct gravure offers precision and color possibilities not otherwise obtainable. The roll of preprinted release is then sent on to the fabric converters for unwinding and pressure-heat lamination of the preprint to the fabric. The fabric is then cooled and rewound and the release sheet discarded.

Furniture. The furniture industry uses a lot of auxiliary release agents in the general gluing and veneering of wood and wood-based materials, as well as in a variety of coating processes. For example, when applying paper impregnated with phenolic and melamine resins for wood surface finishes, the reverse side of the sheet must be coated with a release agent to prevent adhesion to the surface of the press. Mineral oils, paraffin waxes, fatty esters, and specialized paintable silicones are all used for this purpose. Silicone rubber molds are also utilized in the furniture industry, but generally have adequate release properties and require no additional release agents.

Food. Familiarity with household food uses of release agents, such as in polyolefin refrigerator ice trays, gelatin dessert molds, wax paper, release-coated saucepans and bakeware, and nonstick cooking sprays, is nearly universal. One of the oldest methods is the use of flour to release dough from preparation surfaces. Other natural abherent products include confectioners' sugar, rice flour, and rice paper. The use of abherent surfaces in the food industry not only decreases product loss, but also increases the ease with which the equipment is cleaned. Semipermanent release coatings based on silicone resins and polytetrafluoroethylene can last many months before they must be replaced, whereas wax coatings are generally removed after each batch of cooking during cleaning operations and then reapplied prior to the next use. Vegetable and animal fats are still used in the bakery industry in this way and are applied pure or as emulsions. Bakery liners are also employed, and they can often be reused several times. Chromium complex-treated glassine and parchment papers are examples. It should be noted that the chromium in these complexes is trivalent and does not have the hexavalent chromium health hazard.

Economic Aspects

The low surface energy and inertness of most of the materials used as release agents are exploited in numerous other coating and processing aid applications, such as water repellents, polishes, defoamers, and pigment and filler treatments. Consequently, definitive economic information is rarely available. The 1992 worldwide consumption of lubricants and mold-release agents for plastics additives is given as ca 130–140 × 10³ metric tons, with the United States, Western Europe, and the rest of the world accounting for one-third each (25). This breakdown of total worldwide consumption is consistent with an earlier 1982 survey of U.S. plastics additive lubricant consumption of ca 40 × 10³ t. These two estimates include both internal and external lubricant applications. Consumption is equally divided worldwide between fatty acids and esters, metallic soaps, amide waxes, paraffin and polyethylene waxes, and miscellaneous types (25). More detailed U.S. metallic soap production figures have been given (see DRIERS AND METALLIC SOAPS).

Some limited economic data are also available for certain of the industrial applications that are critically dependent on release agents. The pressure-sensitive tape industry is an example. This application is dominated by silicones, and in the United States in 1982 silicones valued at an estimated \$75 × 10⁶ were used for paper coatings for this industry. This estimate is consistent with a later 1988 estimate for worldwide sales of silicone release agents of all types of \$130–150 × 10⁶ (26). Growth is limited in many of these release applications, with price increases largely offset by production efficiencies. Adhesives volume has only grown overall since the mid-1980s at a 2° annual rate (27), and it is likely that the related abhesive application has grown similarly. Release agent product prices vary from ~\$2 to >\$80/kg (3), depending on the type and concentration of active constituent. The lower cost products are simple solvent dispersions of stearates and paraffin waxes. The more expensive products are complex, catalyzed reactive proprietary compositions. They may sometimes be the only alternative, but they are often cost effective even when less expensive materials can be used, because of the lower amount generally required or their multiple-release usability. A release agent should be selected not on the basis of unit cost, but for its impact on overall processing and finishing costs.

Health and Safety Factors

Most general-purpose release agents have been developed for this market in part because of their low toxicity and chemical inertness and do not usually present health and safety problems. Some of the solvent dispersions require appropriate care in handling volatile solvents, and many suppliers are offering water-based alternatives. Some of the solids, particularly finely divided hydrophobic solids, can also present inhalation problems. Some of the metallic soaps are toxic, although there is a trend away from the heavier, more toxic metals such as lead. The reactive type of release coating with monomers, prepolymers, and catalysts often presents specific handling difficulties. The potential user with health and safety questions is advised to consult the manufacturer directly.

Principal health and safety concerns involve contact with foodstuffs and drugs. U.S. government regulations governing the use of additives such as

release agents are listed in the *Code of Federal Regulations*. Title 21 contains the rules established by the U.S. Food and Drug Administration. Part 175 deals with indirect food additives, ie, adhesives and components of coatings. Specifically, Part 175.300 covers resinous and polymeric materials for use as components of coatings. This regulation lists many substances, some of which have a release effect, and it contains a section on release agents that are the basic polymer of the coating, including polyethylene wax, polytetrafluoroethylene, and silicones, as well as a section on surface lubricants, including glyceryl monostearate [31566-31-1]. Regulations are subject to change, and the *Code of Federal Regulations* is revised at least once each calendar year. It is also kept up-to-date by the individual issues of the *Federal Register*, a daily government publication.

Although the intent may be the same, the details of regulatory practice differ in other countries, and a multitude of regulations exists worldwide. Given the changing regulatory climate, the different practices among countries, and the wide range of release agent materials available, health and safety information on a particular product is best acquired directly from the local formulator or manufacturer. Their skill and experience are often the best defense against hazards associated with the use of release agents.

BIBLIOGRAPHY

"Abherents" in *ECT* 2nd ed., Vol. 1, pp. 1–11, by G. P. Kovach, Foster Grant Co.; in *ECT* 3rd ed., Vol. 1, pp. 1–9, by R. D. Cowell and B. R. Bluestein, Witco Chemical Corp.

1. C. J. Reilly, in P. A. Toensmeier, ed., *Modern Plastics Encyclopedia '94*, Vol. 70 (12), McGraw-Hill Book Co., Inc., New York, 1993, p. 233.
2. *Thomas Register of American Manufacturers*, 80th ed., Thomas Publishing Co., New York, 1990.
3. C. Kirkland, *Plast. Technol.* **26**, 65 (1980).
4. *Plast. World*, **41**, 100 (1983).
5. *Annual Book of ASTM Standards*, Vols. 08.01 and 08.02, American Society of Testing and Materials, Philadelphia, Pa., 1993.
6. W. D. Harkins, in J. Alexander, ed., *Colloid Chemistry*, Vol. 1, Chemical Catalog Co., Inc., New York, 1926, p. 227.
7. J. N. Israelachvili, *Intermolecular and Surface Forces*, Academic Press, Inc., San Diego, Calif., 1989, p. 21.
8. W. A. Zisman, *Adv. Chem. Ser.* **43**, 1 (1964).
9. F. E. Bartell and H. H. Zuidema, *J. Am. Chem. Soc.* **58**, 1453 (1936).
10. D. K. Owens and R. C. Wendt, *J. Appl. Polym. Sci.* **13**, 1741 (1969).
11. E. G. Shafrin and W. A. Zisman, *J. Phys. Chem.* **61**, 1046 (1957).
12. E. G. Shafrin, in I. Skeist, ed., *Handbook of Adhesives*, Van Nostrand Reinhold Co., New York, 1977, p. 67.
13. E. G. Shafrin, in J. Brandrup and E. H. Immergut, eds., *Polymer Handbook*, 2nd ed., John Wiley & Sons, Inc., New York, 1975, p. III–221.
14. W. A. Zisman, in P. Weiss, ed., *Symposium on Adhesion and Cohesion*, Elsevier Science, New York, 1962, p. 176.
15. R. J. Roe, *J. Phys. Chem.* **72**, 2013 (1968).
16. H. Kobayashi and M. J. Owen, *Macromols.* **23**, 4929 (1990).
17. P. J. Fagan and co-workers, *Science* **262**, 404 (1993).
18. S. H. Collins, *Plast. Compounding* **6**, 38 (1983).
19. G. L. Schneberger and T. Nakanishi, *Polym. Eng. Sci.* **21**, 381 (1981).
20. J. Stein, T. M. Leonard, and G. A. Smith, *J. Appl. Polym. Sci.* **42**, 2355 (1991).
21. L. A. Duel and M. J. Owen, *J. Adhesion* **16**, 49 (1983).
22. D. B. Cox, *ACS Symp. Ser.* **172**, 565 (1981).
23. B. J. Briscoe and S. S. Panesar, *J. Adhesion Sci. Technol.* **2**, 287 (1988).
24. *Mod. Plast. Int.* **12**, 63 (1982).
25. R. Wolf and B. L. Kaul, in W. Gerhartz, ed., *Ullmann's Encyclopedia of Industrial Chemistry*, Vol. A20, VCH Publishers Inc., New York, 1992, p. 459.
26. R. P. Eckberg, in D. Satas, ed., *Coatings Technology Handbook*, Marcel Dekker, Inc., New York, 1991, p. 621.
27. *Adhesives Age*, 46 (Oct. 1993).

General References

H. Lammerting, "Release Agents," in W. Gerhartz, ed., *Ullmann's Encyclopedia of Industrial Chemistry*, Vol. A23, VCH Publishers Inc., New York, 1993, p. 67.

D. Satas, *Coatings Technology Handbook*, Marcel Dekker, Inc., New York, 1991.

J. Stepek and H. Daoust, *Additives for Plastics*, Springer-Verlag, New York, 1983, p. 34.

M. J. Owen, in J. I. Kroschwitz, ed., *Encyclopedia of Polymer Science and Engineering*, Vol. 14, 2nd ed., Wiley-Interscience, New York, 1988, p. 411.

Michael J. Owen
Dow Corning Corporation

RENEWABLE ENERGY RESOURCES

The ongoing, continuing drive to develop renewable energy resources began in the energy-short 1970s. Sharply rising oil prices and an increasing dependence on international markets for energy supplies served to shift the focus of planners toward those limitless resources that stand available for exploitation and at the same time are virtually without emissions as a result of their operation. By harnessing the sun, wind, falling water, plant matter, and heat from the earth, it was asserted, the environmental impact of energy use would decrease and the nation would regain its energy independence. In this vision for the future, renewable energy resources would complement fossil fuels and, eventually, emerge as a significant energy source.

The promise of such a scenario spurred legislative incentives for the development of renewable energy technologies. One of the most significant of these was the Public Utilities Regulatory Policies Act (PURPA) of 1978, enacted to encourage development of alternative energy sources and to reduce dependence on oil and gas. The enactment of PURPA obligated utilities to purchase the electricity from cogeneration sources and small power producers, ranging from hydroelectric and wind projects to municipal waste combustors. In the years following its passage, the electricity generated by Qualifying Facilities (QF) defined by the legislation generally has tended to replace electricity from a utility system's least efficient plants, typically older coal- and oil-fired facilities.

The incentives for the development of renewable energy and energy-efficient technologies increased with the continued passage of state and federal legislation, particularly the federal Energy Policy Act of 1992. This act encouraged the production and utilization of renewable energy resources, including advanced technologies. Also, exports of U.S. renewable energy technologies and services were promoted.

Despite such incentives, only about 2% of U.S. electricity needs in the mid-1990s are supplied by renewable energy technologies other than hydropower. Most of the nonhydro renewable power comes through some form of combustion, such as the burning of biomass (wood and agricultural waste), landfill gas, or municipal solid waste. Thus, as of the mid-1990s, only a small fraction of the nation's electricity comes from solar, wind, and geothermal sources.

The limitations on renewable energy's progress to date result from its relatively high cost and the economics of the utility industry, which is faced with an increasingly competitive environment and is generally awash in surplus power. Although renewable energy technologies have overcome a number of formidable technical hurdles to bring costs down and increase reliability, further progress has been prevented by the lower cost of natural gas and the efficiency of gas-fired generating plants.

Nevertheless, as costs continue to decline, the renewable energy market can be expected to grow, particularly if possible global warming trends continue to be linked with greenhouse gases such as those emitted by fossil fuels. Other factors that are increasing interest in renewable energy include niche cost advantages, regulatory pressures, customer service requirements, fuel flexibility, and security.

The extent to which each technology is poised to advance is described in separate discussions of photovoltaics, solar-thermal power, and wind, biomass, waste-to-energy, geothermal, hydropower, and wave energy.

Photovoltaic Cells

Since the first photovoltaic (PV) cells were fabricated for the U.S. space program in 1958, PV technology has evolved from once being a very high cost but essential and effective space power source to later becoming a small but diversified and enduring worldwide industry (1–6). Led by firms based in the United States, Japan, and Germany, this industry serves multiple markets (see PHOTOVOLTAIC CELLS).

Since the mid-1980s, PV sales have climbed by a factor of four or more to over 60 MW annually, while installed costs have fallen by more than half. In this same period, a substantial consumer market for PV-powered watches and calculators emerged in Japan. The total number of these products in the marketplace in the 1990s numbers in the hundreds of millions.

Conventional power plants, however, range in size from 100 to 1000 MW. To enter that kind of bulk power market, the lowest quoted costs, ca 1995, for installed PV systems, about \$6–\$7 /W, need to decrease by a factor of three or more to the \$2 /W that is considered competitive with conventional sources of bulk power generation. However, steady progress in advanced PV technologies, such as multilayered thin films and high efficiency concentrator systems, suggests that competitive PV systems may become available before the year 2005. These systems would be capable of generating electricity at 6–8¢ /kWh by offering a net system efficiency of 15% and a system cost of \$60–\$100 /m².

Taken as a group, PV cells comprise solid-state devices in which photons of light collide with atoms and transfer their energy to electrons. These electrons flow into wires that are connected to the cells, thereby providing current to electrical loads.

PV systems consist of arrays of cells that are interconnected in panels or modules to increase total power output. Often the systems include sun-tracking equipment, as well as power-conditioning equipment to convert dc to ac. The systems can range in size from a simple one-panel, fixed-orientation unit to a vast field of modules that accurately track the movement of the sun. Electric utilities in Europe, Japan, and the United States have hosted several experimental PV power plants. The largest to date was a 5.5-MW plant at Carrisa Plains, California, built by Siemens Solar Industries (formerly ARCO Solar).

PV cells may be fabricated using one of four production methods. In the traditional method, which still accounts for a large portion of annual production, PV cells are made from slices of single-crystal silicon. In this process, crystalline silicon is grown in pure ingots made from molten silicon, and is then cut into wafers. The wafers are polished and processed into PV cells, which are mounted in modules. The ingot-based technologies of the 1990s also use polycrystalline silicon.

A second method grows silicon ribbons or sheets directly, bypassing the ingot stage. The sheets are cut into cell-size pieces that are processed to make cells. This manufacturing approach consumes less silicon than ingot-based technologies.

Still another method used to produce PV cells is provided by thin-film technologies. Thin films are made by depositing semiconductor materials on a solid substrate such as glass or metal sheet. Among the wide variety of thin-film materials under development are amorphous silicon, polycrystalline silicon, copper indium diselenide, and cadmium telluride. Additionally, development of multijunction thin-film PV cells is being explored. These cells use multiple layers of thin-film silicon alloys or other semiconductors tailored to respond to specific portions of the light spectrum.

Compared to ingot-based and silicon sheet technologies, thin-film modules require less semiconductor material and can be more highly automated; both attributes lead to lower cost. However, the performance of thin-film modules has yet to equal that of ingot-based and silicon sheet technologies.

A fourth category of PV technologies, concentrator photovoltaics, uses small but very efficient cells illuminated with concentrated sunlight. PV concentrators use lenses or reflective devices and track the sun through daily cycles. The tracking maintains the concentrated light at intensities up to several hundred times normal sunlight.

Most current world PV production comes from the classical ingot technologies. Nevertheless, sheet growth, thin films, and PV concentrators hold greater potential for sufficient cost reduction to stimulate large-scale applications.

The evolution of the competing technologies has been marked by ever-increasing efficiencies in converting sunlight into electricity. The earliest PV cells converted <5% of the sun's energy to electricity. By contrast, recent tandem cells designed for use with highly concentrated sunlight have exceeded 30% efficiency in the laboratory.

In the field, commercial-scale electrical energy conversion from sunlight exceeded 20% in an integrated array system patented by the Electric Power Research Institute (EPRI). The system employs an advanced crystalline silicon cell that uses radiation-hardening technology developed for space satellite applications. The system was built by Amonix, Inc., which is leading a team of EPRI contractors to complete and commercialize the innovation. The pace-setting efficiency was achieved in a 2000-W testbed array installed at the Georgia Power Company (Figs. 1 and 2). Individual solar cells (metallized on the bottom side) (Fig. 2b) span gaps in the copper conductive layer. In the 20-kW array, 168 such panels (120 W per panel), each containing 24 cells, and additional optical elements form the bottom portion of a box-beam structure. The top part of the structure consists of parquets of Fresnel lenses designed for 250 $\times$ sunlight concentration. Made of molded acrylic, the lenses in each parquet are arranged in a 4 $\times$ 6 matrix. A motorized, computer-controlled pedestal keeps the array pointed at the sun. Based on this array, a 20-kW demonstration system is under construction for utility evaluation at an Arizona Public Service facility.

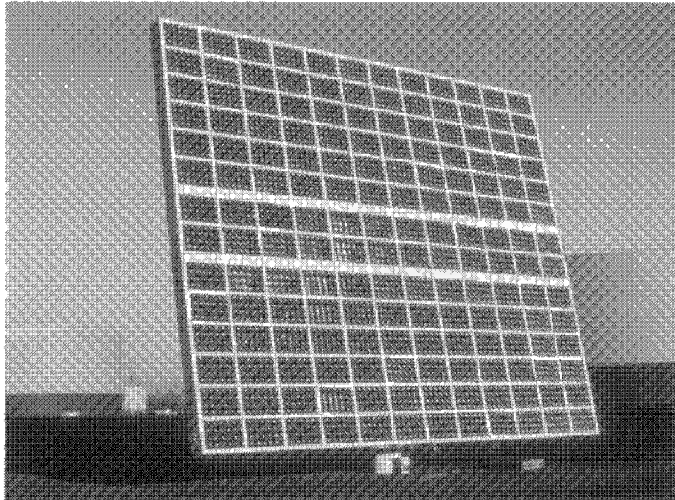


Fig. 1. An integrated high concentration photovoltaic array that achieved a solar conversion efficiency exceeding 20% in a 2000-W testbed (7).

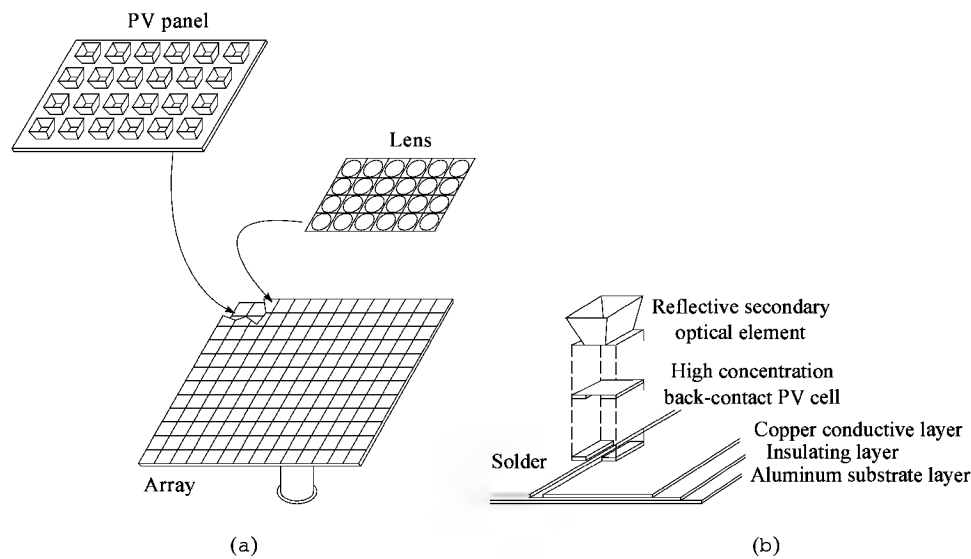


Fig. 2. (a) The cell mount design of the integrated PV array of Figure 1, which uses laminated conductive and insulating layers on top of an aluminium substrate in a printed circuit board-type panel (7). The array produces 20 kW at 20°C ambient and 850 W m² direct sunlight, and measures 155 m². The lens is a molded acrylic Fresnel lens parquet mounted on the front of the array structure. The PV panel is mounted on the back of the array structure and is made up of (b) individual solar cells (7).

In addition, other utilities are installing established solar cells in a growing number of tests that may lead to a mass market. The studies may indicate the extent to which solar cells can be used to avoid installation costs for new distribution lines between conventional power plants and remote customers' buildings. Also, among other objectives, PV cells may provide an economical means of helping to supply demand during peak summer periods in northern climates.

Solar-Thermal Power

The first solar-electric technology to arouse industry interest was solar-thermal energy (1,3,5,6,8). Under favorable circumstances, it can be cost-effective, as evidenced by the fact that solar-thermal gas-hybrid plants produce over 350 MW of commercial power in southern California. This power is used during peak demand to supplement that available from conventional generation.

The level of benefits from tax credits and favorable power purchase terms that helped give these installations their start has diminished in recent years. Nevertheless, with the remaining tax credits available and given current (ca 1995) natural gas prices, solar-thermal technology can deliver power at 8–12¢ kW-h and an installed cost of \$2500–\$3000 kW.

Solar-thermal technology uses tracking mirrors to concentrate sunlight onto a receiver. In turn, the receiver absorbs solar energy as heat, warming a fluid that then drives a turbine generator. Most solar-thermal plants require cooling water.

Solar-thermal receivers may be centralized or distributed. Central receiver systems use fields of tracking mirrors, or heliostats, to focus sunlight onto a tower-mounted receiver (Figs. 3 and 4). Distributed receivers use point-focusing parabolic dishes or line-focusing parabolic troughs to concentrate sunlight. Solarthermal power plants may be entirely solar or they may be hybrids that use fossil fuels to boost power output or extend operating hours.

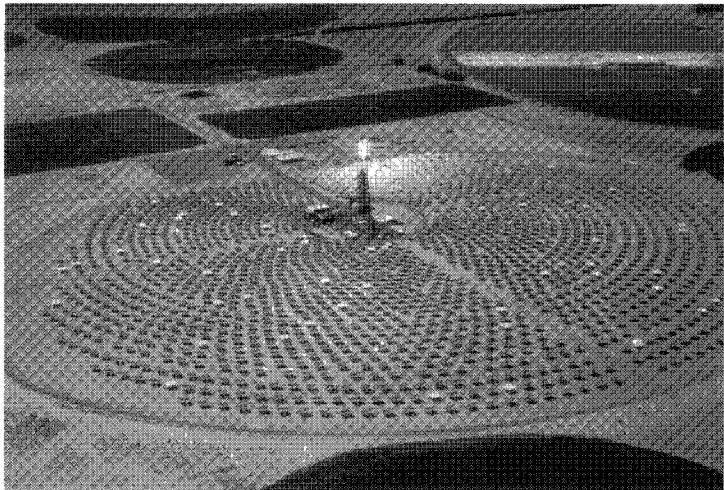


Fig. 3. The 10-MW Solar One plant, which advanced solar thermal power through the use of tracking mirrors to concentrate sunlight onto a central tower-mounted receiver.

Courtesy of Southern California Edison.

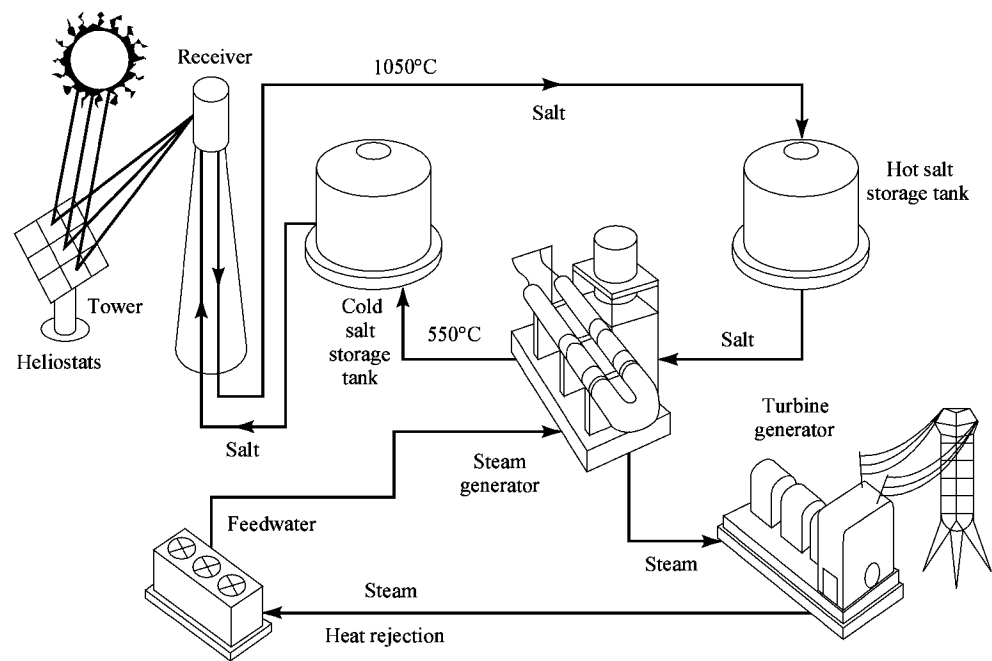


Fig. 4. In the Solar Two Project a molten salt system shown in the scheme replaces Solar One's water steam system. In operation, "cold" molten salt is pumped from a storage tank to a receiver on a tower. Sunlight reflected from a field of sun-tracking mirrors heats the salt in the receiver to 1050°C. The heated salt then flows down into a hot storage tank where it is pumped to a heat exchanger to produce the steam that drives a turbine. Some of the hot molten salt can also be stored to produce steam on demand at a later time. Salt cooled to 550°C in the steam generator recirculates through the system and is reheated in the receiver. Courtesy of Southern California Edison.

The pioneering 10-MW Solar One plant in Barstow, California, produced the most successful central receiver tests. The plant was funded primarily by the U.S. Department of Energy (DOE) and operated by Southern California Edison Company in the early to mid-1980s. EPRI provided technical evaluations of the experiment. The general conclusion drawn from Solar One and other experiments was that further substantial engineering development was needed.

Based on the results of the Solar One plant, Southern California Edison formed a consortium that included DOE and EPRI to construct a Solar Two Project. Solar Two will convert the idle Solar One central receiver plant from a water steam system to a molten salt system, thereby improving efficiency and operating performance. With the molten salt technology, solar energy can be collected during the day and stored in the salt to produce electricity when needed. The three-year demonstration is scheduled to begin in late 1996.

Central receiver systems not only require components that can withstand severe and frequent thermal cycling, but in addition they entail long warmup times and exhibit slow transient responses. As a result, energy production from the best systems have been about half of that expected. As development complexities became apparent, government support was curtailed and industrial commitment waned.

The future of distributed receivers using dish concentrator systems may rest with the development of several components, beginning with a highly efficient, easily maintainable power conversion unit that can convert concentrated solar energy into electricity. Stirling engines, which are being developed for other applications, appear to be the most appropriate choice for this function. If they become commercially successful and sufficient means to support system development efforts materialize, this category of solar-thermal technology may be pursued.

By contrast, trough concentrators are supplying the 350 MW of commercial power. Several plants, all using line-focusing parabolic troughs to concentrate sunlight, were placed into service in the mid- to late 1980s by LUZ International Ltd. The solar-hybrid plants use natural gas to supply 25% of the electricity produced. Assisted by state and federal tax credits and favorable long-term energy purchase agreements with utilities, LUZ shouldered most of the financial risks of development and commercialization. Although the company was unable to continue in business, its operating plants demonstrated that solar energy technology can be made viable.

Passive Solar Power. The largest opportunity for solar power may be in building and process design, among other applications that take advantage of natural lighting and cooling. Such applications will not generate electricity, but will reduce demand for commercially generated electricity. Window technologies and various architectural techniques allow passive solar methods to be readily incorporated into new building designs. Potential heating and cooling energy savings of 15–20% can often be achieved at comparable construction costs. Widespread use of passive solar devices and methods may require energy efficiency standards for building systems.

Wind Power

Like solar-thermal technology, wind is providing utility customers with electricity (3,6,9,10). The technology is advancing despite the expiration of favorable tax credits in the mid-1980s. About 17,000 mid-size wind turbines, nearly all of them in California, are producing approximately 1500 MW of electricity and more are planned. Many of these turbines are operated by Kenetech Corporation, the world's largest developer of wind power. The electricity produced costs 7–9¢ kW·h, and state-of-the-art technology reduces this amount to 5¢ kW·h (constant dollars). However, even this amount exceeds the present cost of wholesale gas of about 2–3¢ kW·h. Total installed costs for state-of-the-art wind systems range from \$900–\$1200 kW.

Wind turbines have had a varied history. Once widely used for electric power generation in remote areas in the United States, for example, they eventually fell into disuse by the 1940s as a result of rural electrification. In the 1970s, interest in wind revived in the face of the energy shortages. Government and industry combined efforts to develop, build, and test over a dozen new turbines, the largest of which was capable of generating several megawatts. However, their imposing size and complexity and the high costs associated with their operation and maintenance discouraged potential commercial developers. Even the turbines on the order of hundreds of kilowatts were not yet cost-competitive with conventional forms of electric generation.

In the early 1980s, favorable tax credits and energy rates for independent power producers encouraged the development of wind farms in California based on 50–100-kW turbines. Simpler and relatively easy to design, build, install, and repair compared to the earlier, larger-size turbines, this new generation of wind machines was also relatively reliable and provided lower cost electricity than the previous generation. Nevertheless, as numerous manufacturers were drawn into the business and wind farms came into existence in California, machines of widely varying effectiveness were deployed. However, enough of the wind farms pulled through to renew utility interest.

Typical wind turbines consist of rotor blades mounted atop a tower and connected by gears to a drive shaft that spins a generator. Another common design is the vertical axis turbine, which has an eggbeater-shaped rotor attached directly to a vertical shaft. The rotor blade length and wind speed determine the amount of electric power that can be delivered. In general, wind speeds of at least 6.7 m s are sought for power generation.

Most of the best locations for wind projects lie outside California. Several northern Rocky Mountain states and Northern Plains states possess substantial resources. Also, the Northeast and Texas have considerable wind resources. In all, about 14 states each possess a wind energy potential that is equal to or greater than that of California.

A cooperative alliance of Kenetech Corporation, Pacific Gas and Electric Company, Niagara Mohawk Power Corporation, and EPRI has developed an advanced, utility-grade variable speed wind turbine that can deliver 300 kW. Commercially available and in widespread use, the turbine uses advanced power electronics to increase turbine efficiency, improve power quality, and lengthen turbine life. Future efforts are directed at developing turbines capable of individually producing 500–1000 kW for eventual deployment in a large-scale power role.

Traditionally, wind turbines have operated at constant rpm to produce 60 Hz a-c power. Because the extra torque generated by wind gusts must be absorbed by the drive trains of constant speed wind turbines, they require heavier designs than comparable variable speed models. By contrast, variable speed turbines employ a power electronic converter between the generator and the utility power line. The converter allows the rotor and generator to speed up with gusts or stronger winds, without increasing drive-train torque. The increased rotational energy is converted into additional electricity. Energy capture increases by 10% or more, and stresses on the turbine are reduced. The variable speed design produces wind power at a cost of 5¢ kW·h, a level that is also achieved in moderate wind resources by other, fixed speed turbines, eg, the AWT-26, Z, and Z-46.

In Europe, government policies (ca 1995) are calling for a steadily increasing commitment to wind power. Combined, the European programs call for the installation of at least 4000 MW by the twenty-first century, a level that would dominate world production. Environmental concerns are the incentives behind Europe's wind targets. With over 2000 MW of wind power already installed, Europe is well on its way to achieving its goal.

Biomass Fuel

Concern over possible global warming trends has been linked with steady increases in greenhouse gases, such as carbon dioxide, CO₂. This gas is emitted whenever fossil fuels or biomass materials, such as wood and agricultural wastes, are burned. When used as a renewable fuel, however, the CO₂ released by biomass during combustion ideally equals that consumed when the fuel was grown. Thus, biomass should not contribute to CO₂-related global climate change (3,6,11).

Global demand for electricity can be expected to eventually increase substantially over the levels of the mid-1990s. In that event, only a massive expansion in the use of biomass and other nonfossil fuel sources can slow the annual increase in global CO₂ emissions.

Also, wood fuel is low in sulfur, ash, and trace toxic metals. Wood-fired power plants emit about 45% less nitrogen oxides, NO_x, than coal-fired units. Legislation intended to reduce sulfur oxides, SO_x, and NO_x emissions may therefore result in the encouragement of wood-burning or cofiring wood with coal.

In the United States, up to about 4 × 10¹⁵ Btu yr of biofuels are consumed for electricity generation, raising process heat, and domestic heat. Furthermore, much of the energy needs of many nations are met by biofuels, including wood and wood waste, spent pulping liquors, bagasse, and municipal waste. Some use is also made of dried corn cobs, rice hulls, and a wide variety of agricultural wastes used in niche applications.

About 6000 MW of electricity generating capacity in the United States is based on the operation of several hundred wood-fired plants. Most of them are owned by paper companies and saw mills, which burn their own scrap wood to generate heat and electricity, primarily for on-site use. Excess electricity is commonly sold to utilities. Fewer than 10 of the country's wood-fired plants, generating less than 300 MW, are actually operated by utilities; other wood-fired plants have been built and are operated by independent producers. The largest units range between 50–60 MW.

Extensive efforts are underway to use wood and agricultural wastes as fuel, including an increasing emphasis on capturing these materials. Also, traditional direct combustion technologies used for electricity generation have been joined by developments in the use of circulating fluidized-bed technology for biofuels, and cofiring as a technology for using biofuels in pulverized coal and cyclone boilers. Existing and emerging gasification technologies are also under study for use in supporting conventional combustion turbine technology or for integrated gasification–combined cycle (IGCC) systems. Experiments continue with direct-fired gas turbines.

The higher utility energy cost of a wood-fired plant, compared with the corresponding cost for a coal-fired plant, arises from the higher costs entailed in collecting and transporting wood and the lower energy heating value of wood relative to coal.

Short-rotation woody crop (SRWC) practices offer one means to reduce wood collection costs. The SRWC programs seek to develop a steadily replenishable, predictable supply of wood for energy production. For example, SRWC plantations, focusing mainly on fast-growing, short-lived hybrid willow trees, use many typical agricultural practices, including fertilizers and pesticides to maximize yields of genetically improved trees. The fastest growing hybrids have produced growth rates exceeding 24.7 dry tons ha yr. The first harvests are expected three to four years after planting. Sponsors for such efforts include the U.S. Department of Energy, EPRI, Empire State Electric Energy Research Corporation, New York State Electric & Gas Corporation, and Niagara Mohawk Power Corporation.

Also, EPRI is investigating whole-tree burn power plants that will dry and burn SRWC fuel without requiring tree trimming, wood chipping, or other fuel preparation. The concept involves the combustion of tree crops from farms distributed within a 40-km radius of a given plant.

Harvested and delivered whole, the trees are dried in an air-supported fiber glass dome structure over a 30-d period by using waste heat from the combustion process in the adjacent plant (Fig. 5). Trees leave the dome on the conveyor and, at the boiler wall, batches are cut into sections to fit the boiler. These sections are about 8.5 m long for the 100-MW facility studied by EPRI and the Minnesota Power & Light Company.

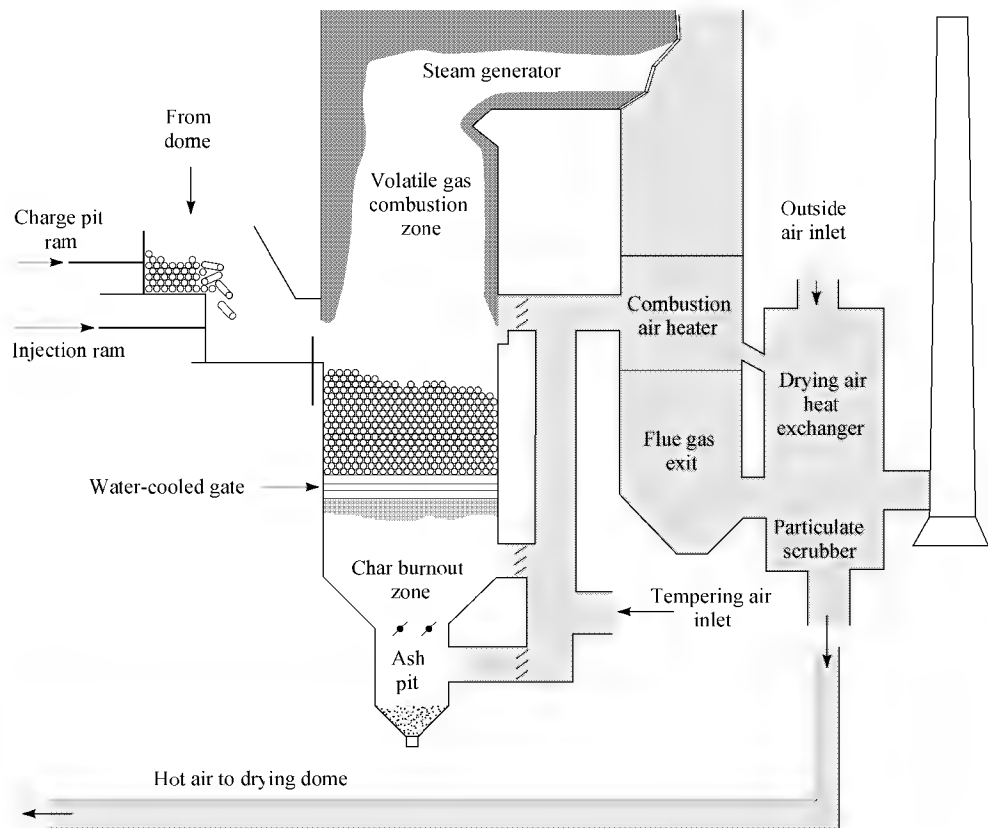


Fig. 5. Developmental concept for a process for the conversion of trees to electricity (12). Dried trees are cut into sections that will fit the boiler chamber. Successive rams push the trees into a charge pit, and then into the furnace. Although larger than a gas-fired boiler, the whole-tree-energy boiler is otherwise similar. The greater height helps achieve a high heat-release rate and complete combustion. Cinders from the burning bed of trees fall through a grate into an area where any remaining carbon in the material burns away. Air is fed into the boiler both below and above the bed of trees to promote complete combustion. Waste heat is fed into an adjacent tree-drying dome.

In theory, whole-tree-energy plants have the potential to be more efficient than existing wood-fired generators, which are fueled by chipped wood with a relatively high moisture content (45% o). The dried whole trees have a moisture content below 25% o, and whole-tree plants potentially can be built to have a greater capacity and to employ high pressure, high temperature steam.

Waste-to-Energy

Large quantities of municipal solid waste (MSW), comprised mostly of residential and light commercial waste, are generated in the United States at the rate of approximately 180×10^6 t yr. Although the composition of MSW varies according to the source and time of year in which it is generated and collected, MSW generally comprises paper, cardboard, wood, plastic, garbage, food and beverage containers, metal, glass, organic material from yards, appliances, and miscellaneous materials, such as rugs, blankets, shoes, mattresses, and telephone wire.

The conventional means of disposing of MSW is by landfilling. About 75% o of MSW is disposed of in this manner, with the balance handled by converting waste to energy (about 15% o) and recycling. However, because landfills are becoming a less acceptable solution, alternative means of disposing MSW have been advanced (Fig. 6).

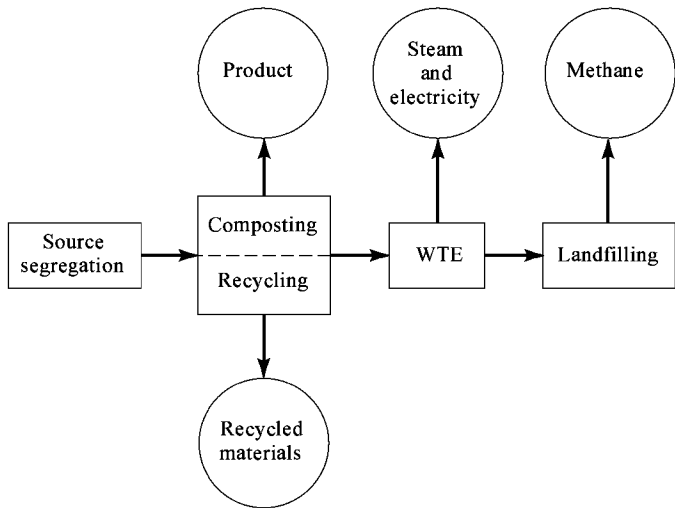


Fig. 6. An integrated approach to the management of municipal solid waste (MSW), advocated by the U.S. EPA, that links source segregation, recycling, waste-to-energy (WTE), and landfilling in a single system. Source segregation refers to the separation of compostable and recyclable components from the balance of the trash at the point where MSW is collected. In source reduction (not shown), another action to reduce waste to landfills, changes are made in the way goods are packaged, distributed, and used (12).

The hazardous components of MSW, ie, household chemicals, oily wastes, and lead and other metals in batteries, can leach from landfills and contaminate both surface water and groundwater or enter the atmosphere. Increased regulation to improve landfill integrity has led to impermeable liners and drainage and water quality monitoring systems. As a result, in many urban areas, land is either no longer readily available for new landfills or is available only at high cost.

Burning of MSW can reduce the volume of landfill space needed by over 80% o. Moreover, MSW represents an available energy source of about

1.5×10^{15} kJ yr (1.4×10^{15} Btu yr) that could provide about 5% of the nation's annual electricity consumption. However, compared to coal, MSW is a poor fuel. Although MSW has a low sulfur content, it may be high in chlorine, aluminum, and some trace metals, including lead and cadmium. The nonhomogeneity of MSW and its high and variable moisture and ash content cause difficulties in maintaining good combustion on a continuous basis.

Nevertheless, plants are disposing of MSW through combustion and recovering energy in the form of steam for electric power generation (3,13). Two commercial technologies are (1) mass burning of MSW as received, using a facility designed for nonuniform, high moisture slow burning materials and (2) processing MSW into refuse-derived fuel (RDF), which is then burned at a cofired or dedicated facility. In cofired plants, RDF and coal are fired simultaneously, whereas at dedicated waste-fired plants, RDF is burned and coal is used only as a backup. Both methods begin by removing unacceptably large items, as well as others, such as discarded gas cans, that might explode during waste processing.

The manufacture of RDF (Fig. 7) entails moving the waste through a series of separators that successively reduce the feed by size and weight; a magnetic separator removes ferrous metals for recycling. Much of the heavier material comprises metal and glass, which can also be recycled. The lighter fraction contains most of the combustible material in the form of a light uniform fluff, RDF, which is conveyed to the power plant for burning. The small residue of unburned waste (ash) can be further treated to remove aggregates and remaining metals before placement in a landfill.

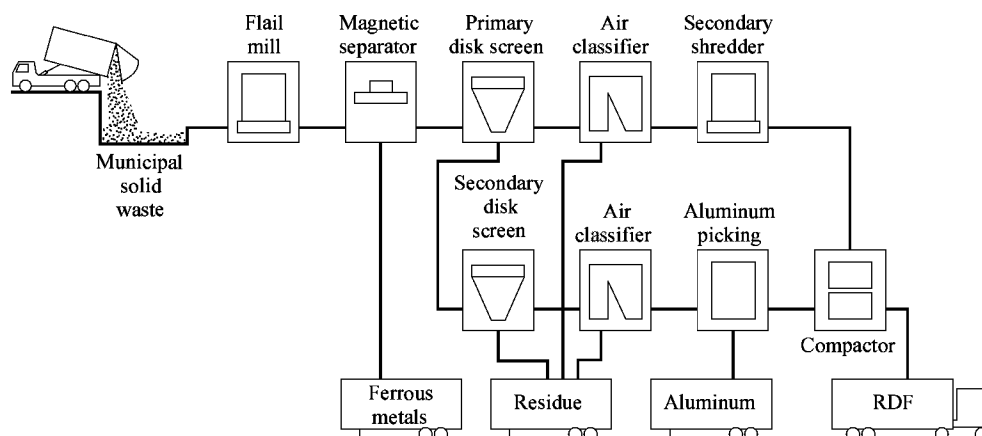


Fig. 7. Scheme demonstrating how MSW can be converted into energy by first processing MSW into refuse-derived fuel (RDF) (12).

An economical means of generating electricity from waste material is to employ a combustor that can burn several different waste materials as fuel. One such option is the fluidized-bed combustor (FBC) which can efficiently burn a variety of inexpensive fuels and comply with environmental regulations. The fuels in the several FBCs in operation in the United States range from RDF, wood, and coal in one to RDF and wastewater sludge filter cake in another. Several FBC units in the construction stage are designed to burn more RDF. In Japan, over 100 FBC units are burning MSW as part of that country's MSW disposal program.

Of the 140 or so waste-to-energy power plants in operation in the United States, over 80% use mass burn, and the balance employ RDF. For the most part, this waste-to-energy (WTE) generating capacity is either owned or being built by organizations other than electric utilities. The utilities own 10 of the 100 plants with an MSW-processing capacity of more than 200 t/d, and purchase energy from 70 others.

By substantially reducing the volume of landfill space needed for MSW disposal, WTE can extend the life of a landfill. In fact, most of the revenues that go to WTE plant operators come from the payments for waste disposal services called tipping fees, and not from electric sales.

Tires. As with MSW disposal, state and local communities have sought increased utility assistance in waste tire management. In the United States, scrap tires are generated at the rate of one tire per person per year, and only 20% are reused or recycled in some fashion. Stockpiles exceed 2×10^9 scrap tires (see RECYCLING, RUBBER).

Landfill operators do not want these tires, and many refuse to accept them or charge high waste disposal fees. Whole tires tend to rise gradually in a landfill, because they fill with methane gas and are less dense than the surrounding moist degrading material. Thus, illegal tire piles abound, posing a fire risk and a health hazard because they make an ideal breeding place for mosquitoes and vermin.

Ironically, scrap tires make good fuel, either whole or as shredded chips, commonly called tire-derived fuel (TDF). Each tire has the heat energy of 3.2×10^6 kJ (300,000 Btu), or about the amount of energy in 13.6 kg (30 lbs) of coal or 9.4 L (2.5 gal) of oil. Also, tires are moderate in both sulfur and ash compared with bituminous coal and do not adversely affect emissions quality.

As an interim measure, burning scrap tires and recovering the energy assists in the solution to the problem of growing mountains of tires, as it reduces the use of nonrenewable fossil fuels. However, the utility industry has long been reluctant to burn materials other than conventional fuels, even though pulp and paper mills and cement kilns have burned mixes of tire chips for many years. In general, the potential savings in fuel costs failed to provide sufficient incentives to offset possible environmental concerns, operational difficulties, and costly equipment modifications. That situation is changing as utilities and other industries are required to adapt to a changing social and economic climate.

Several utilities are burning or have successfully test-burned TDF. For example, the results of a pilot project at Wisconsin Power & Light (WP&L) were so successful that the utility installed its own system to shred tires, thereby assuring a steady supply of uniformly sized tire chips. The tire processing plant will enable the utility to manage about 20% of the 5×10^6 waste tires generated each year in Wisconsin.

The WP&L cyclone boiler will burn TDF continuously with coal, as about 5% of its fuel mix, with little or no modification. By contrast, pulverized-coal boilers, which account for about 80% of the coal-fired capacity in the United States, probably cannot burn tire chips without significant modifications. In these boilers, which burn very fine coal particles in suspension, the heavy chips will fall from the area where best combustion occurs.

Another furnace that does not require fuel preparation is the stoker boiler, which was used by New York State Electric & Gas Corporation (NYSEG) in its TDF tests. At NYSEG, the stoker boiler, which has a 1649°C (3000°F) flame temperature (as does the cyclone boiler), has routinely blended low quality coal, and more recently, wood chips with its standard coal to reduce fuel costs and improve combustion efficiency. In the tire-chip tests, NYSEG burned approximately 1100 t of tire chips (smaller than 5 × 5 cm) mixed with coal and monitored the emissions. The company determined that the emissions were similar to those from burning coal alone. In a second test-burn of 1900 t of TDF, magnetic separation equipment removed metal from the resulting ash, so that it could be recycled as a winter traction agent for roadways.

Landfill Gas Recovery. This process has emerged from the need to better manage landfill operations. Landfill gas is produced naturally: anaerobic bacteria convert the disposed organic matter into methane, carbon monoxide, and other gases. The quantity of methane gas is substantial and could be utilized as fuel, but generally is not. Most of the methane simply leaks into the surrounding atmosphere.

Not all of the gas is wasted. About 300 MW of electricity is generated from landfills. A variety of electric generation systems have been employed by a small number of developers. Most projects use simple technology and are small (2–10 MW). However, an EPRI study has estimated that landfill gas resources in the United States could support 6,000 MW of generation if utilized in 2-MW-sized carbonate fuel cells. Construction on the world's first utility-scale direct carbonate fuel cell demonstration was begun in California. If successful, EPRI estimates that precommercial 3-MW plants based on this design could become available by the end of this decade at an installed cost of \$17,000/kW.

At some landfills, operators have installed flares to combust the gas without recovering any energy. Typically, these cases arise because electricity sell-back rates are too low to justify generation equipment, and laws require a reduction in methane emissions.

Geothermal Power

Although geothermal power generators date back to the early years of the twentieth century, research and development since 1960 have introduced technologies that allow widespread use of geothermal resources of varying quality (14). The principal advances have been in technologies and techniques for exploration, drilling tools, brine handling, power system components, and environmental control. As a result, the installed capacity of geothermal electrical energy worldwide in 1995 reached about 6800 MW (2000 MW in the United States), over 25% of which was achieved since the mid-1980s. Almost 16×10^3 GW·h of geothermal electricity was produced in the United States in 1991.

Because the geologic systems may contain steam or both steam and liquid water, the means to recover energy varies accordingly. If only steam is present, the least common geothermal resource, the steam is fed from the well directly to a steam turbine, which drives an electric generator. If water is also present in a compressed state, a flash-steam cycle may be used: the liquid water flashes into a mixture of hot water and steam by the time the water reaches the lower pressure surface. The steam is then fed to a turbine, and the heat in the hot water may be used elsewhere. Alternatively a newer binary technique uses moderate-temperature geothermal fluids to vaporize low boiling-point organic fluids, which then drive turbogenerators.

The foregoing electric energy is derived from hydrothermal systems, the most commonly used geothermal resource. The systems' hot water or steam are trapped in fractured or porous rocks. Temperatures of the fluids can be as high as 350°C. The high temperature systems used for generating electricity often occur near young volcanoes or where the earth's crust has thinned. Low to moderate temperature systems can be used for direct heating applications.

The most widespread geothermal resource is hot dry rock (HDR), consisting of rocks rich in thermal energy but lacking entrapped water or steam. Harnessing HDR energy would entail drilling a deep well into the hot rock, which is then fractured by the injection of water under high pressure. The fractured zone becomes an engineered reservoir into which surface water is continuously injected and then recovered as hot water or steam. HDR systems are technically feasible, but not yet economical.

Two other localized regions of concentrated heat that are potentially extractable are geopressed geothermal systems and magma. The geopressed geothermal systems comprise hot, high pressure brines containing dissolved methane. Most known geopressed systems are not economical at current (ca 1996) natural gas prices. Pressures can reach 142 MPa (1400 atm). Magma resources comprise partially or completely molten rock, such as may be found in the vicinity of relatively recent volcanic activity. The high temperature of magma, above 650°C, conceivably could be used for producing electric power or high temperature industrial process waste. However, producing magma energy is not yet technically feasible.

Geothermal Heat Pumps. Also called ground-source heat pumps, geothermal heat pumps (GHPs) use the earth as a heat source for heating or as a heat sink for cooling (Fig. 8). This arrangement reduces electricity consumption by about 30% compared to air-source heat pumps. Because of this capability, many U.S. electric utilities and rural cooperatives have promoted the use of GHPs through marketing programs and financial incentives. The result has been that about 200,000 GHPs supply 3,200 GW·h of thermal energy annually to residences and commercial buildings throughout the nation, as of the mid-1990s.

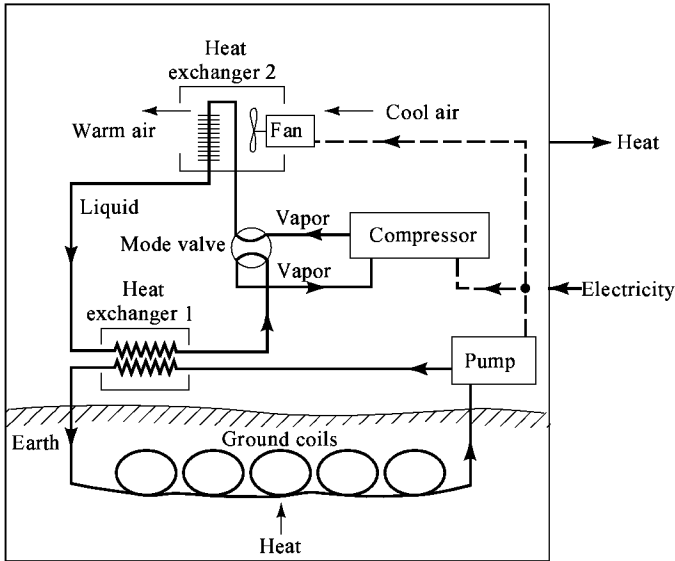


Fig. 8. Geothermal heat pump (GHP). This pump uses the earth as a heat source for heating or as a heat sink for cooling. A water and antifreeze mixture circulates through a pipe buried in the ground (vertically or horizontally) and transfers thermal energy to a heat exchanger in the heat pump. The heat exchanger works through a water-to-refrigerant loop. In a typical reversible heat pump, the ground loop heat exchanger rejects heat from the condenser or delivers heat to the evaporator, depending on the mode of operation.

Courtesy of Princeton Economic Research, Inc. (Rockville, Maryland).

Hydropower

Falling water has been used to generate electricity for over 100 years (3,15). The first hydroelectric (or hydro) power plant was built at Niagara Falls in 1879. In succeeding years, hydro plants were built at other natural waterfalls, as well as artificial ones created by dams, at most of the nation's best sites. In the mid-1990s, hydropower energy accounts for about 93,000 MW or about 12% of the electric generating capacity in the United States bulk power market. The United States also imports some hydropower from Canada. Moreover, the average generating cost of hydropower is less than 3¢ kW·h.

Hydropower provides an essential contribution to the national power grid: its capability to respond in seconds to large and rapidly varying loads, which other baseload plants with steam systems powered by combustion or nuclear processes cannot accommodate. Also, ownership is spread over a broad base. The owners comprise federal and state agencies, cities, metropolitan water districts, irrigation companies, and public and independent utilities. Individual persons also own small plants at remote sites for their own energy needs and for sale to utilities.

The amount of electricity that can be generated at a hydro plant is determined by two factors: head and flow. Head is the distance in elevation between the highest level of the damned water to the point where it goes through the power-producing turbine. Conventional hydro plants must have a head of water that is at least 3 m high to provide sufficient water pressure to operate the turbine. Flow is the rate of water moving through the system. In general, a high head plant needs less water flow than a low head plant to produce the same amount of electricity.

Some hydro plants use pumped storage systems, among the most reliable energy storage systems available. Pumped storage systems use recycled water instead of tapping free-flowing water. After flowing through the turbine, the water resource is pumped, usually through a reversible turbine, from a lower reservoir back to an upper reservoir. Whereas pumped storage facilities are net energy consumers, ie, more energy in total is required for pumping than is generated by the plant, they are valuable to a utility because they operate in a peak-power production mode, when electricity is most costly to produce. The pumping to replenish the upper reservoir is performed during off-peak hours using the utility's least costly resources.

Because the best sites for hydropower dams have already been developed, construction of additional large, conventional plants is unlikely. However,

existing projects could be modified to provide additional generating capacity. Also, many existing dams not equipped for electricity production (only about 3% of the nation's 80,000 dams are used to generate power) could be outfitted with generating capacity. Other opportunities for development are offered by small or low head (from 3 to 9 m) hydro plants at new sites. These concepts have helped to tap vast hydropower resources.

Wave Energy

Ocean waves are formed by the wind driving water toward shore. The wave energy depends strongly on wind speed; the energy is a fifth-power function of speed. Most methods to convert this irregular and oscillating low frequency energy source to grid power employ pneumatic, hydraulic, or hydropower technology (16).

Pneumatic systems use the wave motion to pressurize air in an oscillating water column (OWC). The pressurized air is then passed through an air turbine to generate electricity. In hydraulic systems, wave motion is used to pressurize water or other fluids, which are subsequently passed through a turbine or motor that drives a generator. Hydropower systems concentrate wave peaks and store the water delivered in the waves in an elevated basin. The potential energy supplied runs a low head hydro plant with seawater.

The world's total capacity of grid-connected electric power derived from wave energy is less than half a megawatt, distributed among several demonstration plants. The largest unit, the 350-kWe Tapered Channel plant in Norway, uses the hydropower approach. The plant was developed by Norwave AS and has operated continuously since 1986. Based on this durability, two commercial orders were placed from other parts of the world.

In the United States, other experimental wave-energy systems have been investigated in California and Hawaii, including a 30-MWe heaving-buoy design.

Tidal Power. Tidal power is caused by the gravitational pull of the sun and especially the moon, as they pull at the earth. Reacting to this pull, the ocean's waters rise, causing a high tide where the moon is closest. The difference between low and high tide can range from a few cm to several meters. Harnessing tidal power for electricity production by the use of dams requires a tidal difference of at least 4.5 m, a requirement met at few locations in the United States. Thus, the principal demonstration sites of tidal power are in Canada, China, and France.

Information for this article came from numerous sources, including F. E. Porretto, D. M. DiMeo, and S. Collins (Empire State Electric Energy Research Corp.), E. A. DeMeo and E. E. Hughes (Electric Power Research Institute); J. M. Cohen and D. Entingh (Princeton Economic Research, Inc.); and G. Sommers and J. Renner (Idaho National Engineering Laboratory).

BIBLIOGRAPHY

1.

C. J. Winter, R. L. Sizman, L. I. Vant-Hull, eds., *Solar Power Plants: Fundamentals, Technology, Systems, Economics*, Springer-Verlag, New York, 1991.

2.

T. Moore, *EPRI J.*, 16–25 (Dec. 1992).

3.

R. Golob and E. Brus, *The Almanac of Renewable Energy*, World Information Systems, Henry Holt and Co., New York, 1993.

4.

T. Moore, *EPRI J.*, 6–15 (Oct.–Nov. 1994).

5.

T. Markvart, ed., *Solar Electricity*, John Wiley & Sons, Inc., New York, 1994.

6.

L. Lamarre, *EPRI J.*, 16–25 (Nov.–Dec. 1995).

7.

Integrated High Concentration Photovoltaic Technology, EPRI TR-103267, Electric Power Research Institute, Palo Alto, Calif., 1993.

8.

T. R. Mancini, J. M. Chavez, and G. J. Kolb, *Mech. Eng.*, 74–79 (Aug. 1994).

9.

L. Lamarre, *EPRI J.*, 4–15 (Dec. 1992).

10.

J. Jayadev, *IEEE Spectrum*, 78–83 (Nov. 1995).

11.

L. Lamarre, *EPRI J.*, 16–24 (Jan.–Feb. 1994).

12.

Waste-to-Energy—An Opportunity for Utilities, EPRI TB.65.127.5.91, Electric Power Research Institute, Palo Alto, Calif., 1991.

13.

EPRI J., 28–34 (Sept.–Oct. 1995).

14.

E. Easwaran, D. Entingh, and D. Diachok, *United States Geothermal Technology: Equipment and Services for Worldwide Applications*, DOE EE-0044, U.S. Department of Energy, Washington, D.C., 1995.

15.

DOE Hydropower Program Biennial Report 1992–1993, DOE ID-10424, Idaho National Engineering Laboratory, Idaho Falls, July 1993.

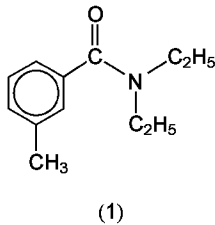
16.

R. J. Seymour, ed., *Ocean Energy Recovery: The State of the Art*, American Society of Civil Engineers, New York, 1992.

Edward A. Torrero
Empire State Electric Energy Research Corporation

REPELLENTS

Repellents are materials that affect insects and other organisms by disrupting their natural behavior of bloodseeking through biting of humans and animals, and are the first line of defense that can be readily used for this purpose (1). The best overall standard repellent is *N,N*-diethyl-*m*-toluamide [134-62-3] (DEET), systematically named *N,N*-diethyl-3-methylbenzamide (1).



Many other compounds are presently in use; a 1993 database search showed 27 active ingredients in 212 products registered by the U.S. EPA for human use as repellents or feeding depressants, including octyl bicycloheptene dicarboxamide (*N*-2-ethylhexyl bicyclo[2.2.1]-5-hepten-2,3-dicarboxamide), dipropyl isocinchomeronate (2,5-pyridine dicarboxylic acid, dipropyl ester), dimethyl phthalate, oil of citronella, cedarwood oil, pyrethrins, and pine tar oil (2). Repellent–toxicant or biting depressant systems are available which are reasonably comfortable for the user and can protect completely against a number of pests for an extended period of time (2).

Newer, cosmetically appealing formulations of chemical repellents have become popular in the United States since the mid-1980s, and interest in personal protection against biting arthropods has been renewed (3). Children and adults may be exposed daily to risks of Lyme disease, and those exposed should be particularly interested in reducing exposure to nymphs of the deer tick *Ixodes scapularis*. The tick, which carries the bacterium *Borellia burgdorferi*, is the same tick species previously called *I. dammini* by some entomologists. Destruction of ticks by spraying is expensive, works only locally, and is not very effective on a large scale. Not only is this disease difficult to diagnose and treat, but treatment of Lyme disease using expensive antibiotics has no effect on the next infective bite. In 1993, the Center for Disease Control (CDC) in Atlanta reported 9677 cases, with the foci in downstate New York and eastern Long Island (ca 3000 per year each). Although most cases are reported around Long Island Sound and in Wisconsin, all U.S. states except one have reported cases. The number of cases reported each year has been relatively constant since 1988, comprising 90% of all vector-borne diseases in the United States. Europe has also reported increased numbers of cases.

An exotic mosquito species, *Aedes albopictus*, the Asian tiger mosquito, has become established in many regions of the United States because viable eggs have been imported in shipments of used automobile tires from Asia that were placed in huge tire dumps all over the country and in the Caribbean islands. This species is a persistent, small but aggressive biter whose females bite any time of the day, are easily disturbed, and cause multiple bites and allergic welts. Although the *Aedes aegypti* yellow-fever mosquito has been displaced from its habitat in Florida, it is not known whether this is helpful for disease transmission, because *A. albopictus* is also an effective carrier of the disease dengue fever. In addition, there are at least 170 other species of mosquitoes in North America, and repellents have been formally tested against only a few regarded as disease vectors (3).

The use of insecticides has led to the rise of widespread resistance in areas where residual insecticides were applied for malaria eradication (see also INSECT CONTROL TECHNOLOGY), and in some locations that have widespread irrigation, as in California, or that practice flooding for rice production. Mosquito resistance to insecticides is prevalent in Southeast Asia, India, and East Africa, as is resistance of the malaria parasites, *Plasmodium falciparum*, also *P. malariae* and *P. vivax*, to prophylactic drugs. The number of cases of malaria is estimated at more than 250 million cases per year, resulting in as many as one million deaths; the goal of malaria eradication has been abandoned except in selected locations (4). The number of cases has stayed rather constant in the 1990s, but with increased danger of mortality on account of the declining effectiveness of drugs. Use of personal protection and repellents is recommended for travelers to malarious areas for avoiding serious health risks.

Disease cases reportable to the CDC of arthropod-borne diseases from 1989 to 1993 appear in Table 1. Arthropod-borne diseases not reportable are babesiosis, tularemia, relapsing fever, Colorado tick fever, leishmaniasis, ehrlichiosis, and LaCrosse virus. Also dengue fever, endemic in the islands of the Caribbean, is readily vectored by *Aedes aegypti* and *Aedes albopictus* (5). Worldwide disease-bearing mosquitoes that are day- or night-biters include the *Aedes* species, which often transmit viruses to humans. Of 435 arboviruses (arthropod-borne viruses) found in insects by 1980, 100 cause diseases in humans. The diseases include yellow fever, dengue haemorrhagic fever, and many forms of encephalitis, including Japanese B, Venezuelan, West Nile, Chikungunya, Ross River, and Rift Valley fever. More than 60 species of *Anopheles* mosquito are vectors of malaria, in particular the human-biting members of the *A. gambiae* species complex in Africa. The *Culex* species are vectors of several forms of encephalitis in the United States, including Western Equine, Eastern Equine, St. Louis, and California encephalitis. Other insect species that can be vectors of diseases are listed in Table 2.

Table 1. Anthropod-Borne Diseases in the United States, 1989–1993

Disease	Number of reported cases
malaria	2400 ^a
encephalitis	1300
encephalitis Calif. serotype	100
sylvatic plague	10
Rocky Mountain spotted fever	100

^a All but eight cases were imported.

Table 2. Insect Vectors of Disease ^a

Insect	Scientific name	Disease
biting midges	<i>Ceratopogonidae</i>	bluetongue virus
blackflies	<i>Simulium</i>	river blindness

sandflies	<i>Phlebotamus, Leptoconops</i>	leishmaniasis
tsetse flies	<i>Glossina</i>	African trypanosomiasis or sleeping sickness
tabanids or horseflies	<i>Tabanus</i>	
clegs	<i>Haemotopota</i>	
deerflies	<i>Chrysops</i>	
snipe flies and stableflies	<i>Stomoxys</i>	
cattle ticks	<i>Rhipicephalus</i>	heartwater fever and blackwater fever
soft ticks	<i>Dermacentor, Amblyomma</i>	Rocky Mountain and spotted fever
mites or chigger mites	<i>Trombididae</i>	scrub typhus
bedbugs	<i>Cimex</i>	
lice	<i>Pediculus</i>	typhus
kissing bugs	<i>Pentatomid</i>	Chagas disease
fleas	<i>Xenopsylla, Pulex</i>	plague

^a Ref. 2.

It has been suggested that the U.S. military, during operations in southeast Asia in the 1960s and 1970s, evacuated more personnel because of vector-borne diseases, primarily malaria, than for combat injuries. Most of the malaria cases reported in the United States since 1990 are imported cases, but there are a few unexplained transmissions. Computer models indicate that disease transmission from mosquitoes in developed countries becomes less likely when there has been a change to Western lifestyles, such as the regular installation of window screens, air conditioning in houses and businesses, and prevalence of indoor television viewing. Thus, the problem of mosquito-borne disease in developed countries may be less serious than in Third World locations where these factors are unlikely to be encountered, and where serious arthropod-borne diseases are common.

Evaluation of Repellents

A critical review (6) of techniques for the evaluation of insect repellents describes many test methods, including the following.

Repellents on Skin. The candidate chemical is dissolved in ethanol and spread over one forearm of the human subject, as DEET (1) is similarly applied to the other forearm. Each arm is then exposed to 1500 avid *A. aegypti* female mosquitoes for 3 min at 30-min intervals. Effectiveness is based on complete protection, ie, the time until the first confirmed bite (one bite followed by another within 30 min).

Repellents on Cloth. Each candidate repellent is applied to a knit cotton stocking or cloth patch at 3.3 g m² cloth, usually as a 1% solution of active ingredient (AI) in acetone. Two hours later, the stock or cloth patch is placed over an untreated nylon stocking on the arm of a subject, the hand covered, and the arm exposed to 1500 female mosquitoes for one minute. If fewer than five bites are counted, the test is repeated at 24 h, then weekly until failure, which is, by definition, five bites per minute. The standard mosquitoes used are *Aedes aegypti*, *Anopheles quadrimaculatus*, or *A. albimanus*. Candidate repellents in cloth tests are in one of the following classes: class 1, effective 0 d; class 2, 1–5 d; class 3, 6–10 d; class 4, 11–21 d; and class 5, >21 d.

Space-Borne Repellents. Air is drawn over a human arm through a 9.5-cm disk of cotton netting having 0.64-cm holes and treated with a solution of the candidate repellent; air is then drawn into an olfactometer cage containing 125 avid female *A. aegypti*. The number of days the repellent prevents >10% of the mosquitoes from passing through the netting constitutes effectiveness. Tests typically use 1.0–3.89 mg g netting.

Skin Patch-Tested Repellents. Small areas of human forearms are marked and treated with small amounts of repellent on a unit area basis to ensure that the treatment rate is always the same between subjects (7). The patches are tested at 0 and 4 hours against small numbers (ca 15) of mosquitoes. This method does not consider creep, movement of repellent across the skin surface, or the interaction between two chemicals owing to such lateral movement of chemical.

Repellents Not Using Human Bait (No Attractant). A treated strip of fabric and a control strip are lowered into a container of crawling arthropods such as ticks, fleas, and mites. After a predetermined time, the strips are lifted, the animals remaining are counted, and the percentage repellency is determined.

Repellents Tested with an Inanimate Attractant. Machines have been constructed by several groups to measure the intrinsic (initial) repellency of a compound when it is added to a warm, moist airstream to overcome the attractiveness of the airstream to mosquitoes. Such machines remove the factor of human odor in attempts to simplify the measurement of repellency.

Repellents Tested with Animal Attractants. Numerous methods have involved the use of animals as attractants, followed by evaluation of repellents as skin treatments or attached cloth treatments, often against crawling arthropods such as fleas, ticks, and mites. Animals such as gerbils, guinea pigs, camels, mice, shaved rabbits, and hairless dogs have been used, particularly when the toxicity is unknown.

Arthropod Repellents

MOSQUITO REPELLENTS

Insecticide-Impregnated Cloth. Clothing impregnants may be applied to socks, head nets, gloves, or conventional apparel by dipping into emulsions, manual application of liquid, or surface spraying with aerosols (qv). Repellents may be applied to cotton cloth using technical material dissolved in acetone, although this solvent may be unavailable or incompatible with some synthetic fabric blends. Repellents may be applied to clothing on a large scale using 10% of a nonionic emulsifier and 90% technical material at 5 wt % of the cloth. Repellent jackets made of 0.64-cm cotton mesh are impregnated with a repellent such as DEET (1). There are several advantages in this. For instance, the repellent is applied to the mesh, not directly on skin. Repellents other than DEET may be used, such as repellents more effective than DEET, against biting flies or other mosquitoes. The jacket may be retreated as necessary simply by the addition of a measured quantity of concentrated repellent, such as 75 wt % DEET, to the jacket while in its storage pouch. An effective skin repellent such as DEET is lost much more slowly from the mesh jacket than from skin. This jacket has been added to the army supply system and is meant to be retreated with standard military-issue 75 wt % DEET. The DEET-treated jacket has been effective against mosquitoes in northern and southern states, blackflies in Maine, biting midges in Florida, and sandflies in Panama. The jacket is effective against heavy pressures of small insects that are capable of passing through mesh. Also, the jacket's wide spaces allow passage of air and moisture, making it a more comfortable alternative than either wearing long-sleeved, heavily repellent-treated clothing or having repellents applied to skin, particularly in warm, humid climates with heavy insect populations (8).

The standard clothing impregnant adapted by the U.S. Armed Services, M-1960, was intended for large-scale utilization in military laundries for protection from mosquitoes, fleas, ticks, and chigger mites (9). It contains equal parts of benzyl benzoate [120-51-4], N-butylacetanilide [91-49-6], 2-ethyl-2-butyl-1,3-propanediol [115-84-4], and a third of a part of Tween 80. The repellent also has disadvantages, including odor, and is less effective as a mosquito repellent than DEET, although benzyl benzoate is a good tick and mite repellent. The M-1960 formulation has been replaced in the U.S. Army system with aerosol formulation DEET for personal protection and clothing treatment or 75 wt % DEET and 25 wt % ethanol in plastic squeeze bottles especially for skin treatment. A formulation problem with aerosols is that chlorofluorocarbon propellants were first replaced with compressed flammable hydrocarbon gases; in the 1990s, however, water-base sprays are more common. Clothing treatments onto 100% cotton military uniforms using DEET have been shown to resist laundering better than treatments onto polyester–cotton blends, whereas 100% polyester retains repellents poorly (10).

Several studies of cloth treatment for toxicity and repellency have been conducted using impregnation with permethrin emulsifiable concentrate (EC), together with residue studies. Permethrin-treated military uniforms, treated by aerosol spray, pressurized spray, or impregnation during the washing process, have been found effective in tests (11). Tent material may be treated with permethrin or dimethyl phthalate for repelling mosquitoes (12). Alternatively, a cloth manufacturer (Graniteville Company) coats bulk cloth with permethrin either before or after the weaving operation. The effective treatment level of 0.125 mg cm² of permethrin that is toxic to mosquitoes appears to be compatible with subsequent sewing operations, thus making

unnecessary post-manufacture or field treatment. This material is commercially available for manufacture into tent fabric and ground cloths.

Bednets and Treated Bednets. Effective personal protection for sleeping can be provided by the use of bednets, of which many varieties are commercially available, including modern self-supported structures. Good information sources are available (2,13). Studies suggest that chemical treatment of bednets at 2 g m² active ingredient can reduce numbers of biting insects in the near vicinity of the bednet, even if the net is torn (14). Mosquitoes that contact a permethrin-toxicant-treated bednet may be killed if they rest on it or may be knocked down and therefore much less likely to bite (15).

Standard Mosquito Repellents. Since its initial report as a promising repellent in 1954, DEET has been considered the best all-around repellent having generally acceptable characteristics, despite a continuing search for a superior chemical. Improvements include many commercial products with added cosmetic agents that use slow release technology, such as the U.S. Armed Services slow release 35% DEET formulation (16). There were 35 EPA-registered repellent products in 1994 that contained only DEET under different trade names (2). DEET is present in 192 of the 212 products mentioned previously (2).

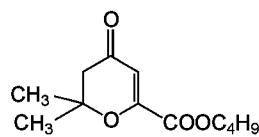
The mode of action of successful repellents is still unproven as of this writing (1996). Replacement of DEET with something better becomes even more remote because of the estimated \$25 million cost of development and registration of a new material. DEET is the principal or only component in most of the effective mosquito repellents on the market, as can be seen from the commercial products listed in Table 3 (2). Muskol (100% DEET) and a number of other products containing 100% DEET were popular in the 1980s, with reports in outdoor magazines and *Consumer Reports* claiming that 100% DEET was perhaps more desirable, and had to be applied less often than more dilute formulations. However, the use of 100% DEET on skin is usually not justified because of the possibility of skin irritation, and not for reasons of better or worse repellent activity (2). Sales of DEET repellents were estimated at \$50–\$75 million yr in the United States in 1984 (17). DEET costs ca \$9 L (\$34 gal) from a chemical manufacturer. A 30-cm³ (1-oz) bottle that retails for \$4–\$5 and lasts the typical consumer a year contains \$0.30 of chemical. More recent tests against *Anopheles albimanus* show good activity (2).

Dimethyl phthalate [131-11-3] (DMP) is a clear oil insoluble in water and soluble in organic solvents, and is synthesized from phthalic acid esterified with methanol. DMP is an effective repellent, used as a standard against *A. aegypti*, and is effective for 11–22 d on cloth. The famous repellent Rutgers 6–12, 2-ethyl-1,3-hexanediol [94-96-2], is a clear oil that is insoluble in water and miscible with organic solvents such as ethanol. Screened during World War II, this repellent is exceptionally effective against *A. aegypti*, lasting 196 d on cloth. Tests have been run against the newer pests *A. albopictus* and *A. aegypti*, including five repellents containing DEET (test standard), a controlled release formulation containing DEET, two dosages of DEET in ethanol, and Avon Skin-So-Soft. On the skin, the repellent chemicals provide significant protection from biting; however, *A. albopictus* is more sensitive to repellents than *A. aegypti*. Two experimental repellents provide 7-h protection, 25% DEET provides >8 h, a controlled release formulation with 35% DEET provides >10 h, but the Avon Skin-So-Soft provides only 0.6-h protection from bites. Permethrin-treated fabric provides complete protection through five washings. Repellent products containing more than 12% DEET can provide satisfactory protection against bites of this species (18). More recently, populations of the Asian tiger mosquito in Texas have been found to show malathion resistance and a tolerance for bendiocarb, which suggests that chemical management of this mosquito may prove to be difficult.

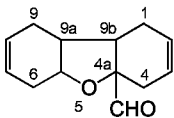
Table 3. U.S. Mosquito Repellents Containing DEET

Company	Location	Product brand name	Form	Composition, %
AMREP, Inc.	Marietta, Ga.	Misty Insect Repellent	spray	7
		Misty Extra Strength	spray	25
ARI, Inc.	Griffin, Ga.	Bug Barrier 100%	lotion	100
		Bug Barrier	spray	7
		Bug Barrier II	spray	25
Bayer Inc.	Chicago, Ill.	Cutter Evergreen Scent	cream	35
		Cutter Insect Repellent	stick	33
		Cutter Insect Repellent	spray	18
		Cutter Maximum Strength Formula 100	lotion	100
		Cutter for Kids	spray	8
CCL Custom Manufacturing Inc.	Danville, Ill.	Personal Insect Repellent	spray	10.5
		CCL Quick Breaking Insect Repellent	foam	8.9
D-Con Co. Inc.	Montvale, N.J.	6-12 Super Strength Premium	spray	26.3
		6-12 Plus Sportsmate II	spray	5
		Premium	cream	45
		Insect Repellent	gel	7
Fuller Brush Co.	Great Bend, Kans.	Ful-Scat Insect Repellent	spray	8
		Adios Insect Repellent	spray	7
Hysan Corp.	Chicago, Ill.	Littlepoint Insect Repellent	lotion	10
Littlepoint Corp.	Cambridge, Mass.	One + One	lotion	11.2
Olson Outdoor Laboratory	Freehold, N.J.	Ole Time Woodsman Kampers	lotion	50
Pete Rickard Inc.	Cobleskill, N.Y.	Ole Time Woodsman Jungle Formula	spray	60
		Insect Repellent Spray for Personal Use	spray	7
Reckitt & Coleman Household Products S.C. Johnson & Son Inc.	Wayne, N.J. Racine, Wis.	Off! Insect Repellent	spray	15
		Maximum Strength Deep Woods Off! 100	lotion	100
		Off! Towelette	wipe	25
		Off! Skintastic II	lotion	7.5
		Unscented Off!	spray	15
		Muskol Insect Repellent	lotion	100
		Muskol Insect Repellent	spray	25
		P P Outdoor Lotion	lotion	19.4
Schering-Plough Health Care Products Inc. Whitmire Research Laboratories, Inc.	St. Louis, Mo.	Whitmire Insect Repellent Stick No. 1	stick	14
		Repel Insect Repellent Towelette	wipe	55
Wisconsin Pharmacal Co.	Jackson, Wis.	Repel 100	lotion	100
		Repel Scented Family Formula	spray	35
		Repel Insect Aerosol	spray	55
		Eddie Bauer Insect Formula	lotion	10
Winsol Labs	Seattle, Wash.	Woodlands Insect Repellent	lotion	10
Woodland Products	Ormond Beach, Fla.			

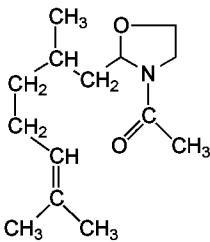
Experimental Mosquito Repellents. A comparison of nine commercial repellents has been made against *A. aegypti* using the skin patch test. At zero time (measuring intrinsic repellency), Stabilene [9003-13-8] and MGK Repellent 326 [136-45-8, 3737-22-2] are significantly inferior to DEET, dibutyl phthalate [84-74-2], indalone [532-34-3] (2), dimethyl phthalate, MGK Repellent 11 [126-15-8] (3), 2-ethyl-1,3-hexanediol [94-96-2], and citronellal [39785-81-4] (4). Efficacy ranking by 4-h ED50 (50% biting rate on the treatment) was indalone, citronellal, 2-ethyl-1,3-hexanediol, and DEET, as the others had become ineffective. The relative superiority of DEET in comparison to other standard repellents has been discussed (19).



(2)



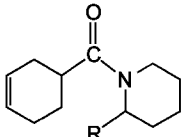
(3)



(4)

Seventy-one *N,N*-dimethylcarboxamides have been tested on cloth against three species of mosquitoes: *A. aegypti*, *A. quadrimaculatus*, and *A. albimanus*. The amides are prepared via the acid chloride by treatment of the appropriate carboxylic acid with thionyl chloride, which is then added to a cold mixture of ether–water containing excess *N,N*-dimethylamine plus sodium hydroxide. Extraction and vacuum distillation give materials that are used as distilled, with 95–98° o purity. Nine of 38 *N,N*-dimethylbenzamides, three of 17 *N,N*-dimethylphenylacetamides, two *N,N*-dimethylphenylcarboxamides, and three *N,N*-dimethylalkanamides are class 5 repellents against all three species of mosquitoes. Typically, *N,N*-dimethylundecanamide [6225-09-8] provided 71-d, 132-d, and 59-d protection against *A. aegypti*, *A. quadrimaculatus*, and *A. albimanus*, respectively (20). Several of these materials have been tested and found to be effective against blackflies in Maine.

Excellent personal protection is afforded by controlled-release topical repellents and permethrin-treated clothing against natural populations of *Aedes taeniorhynchus* (21). Again the increased protection is afforded by use of skin repellents with clothing treatment to avoid all bites. Two of the substituted piperidines, 1-(3-cyclohexen-1-ylcarbonyl)-2-methylpiperidine (A13-37220) [69462-43-7] (5, R = CH₃) and 1-(3-cyclohexen-1-ylcarbonyl)piperidine (A13-22872) [52736-58-0] (5, R = H), have been the most promising in tests against mosquitoes; this piperidine family looks like the only promising group to be discovered in some years (18).



(5)

Tests against *Aedes albopictus* mosquitoes using either piperidine compound have shown these compounds to be nearly as good as DEET. Subsequent tests in southeast Asia suggest repellency about the same as DEET against the resistant *A. dirus* malaria vector species, but poor repellency against *A. albimanus* and *A. quadrimaculatus*; against *A. taeniorhynchus* it shows activity about equal to DEET (22). Further tests of toxicological properties are pending for these two compounds but are suspended at U.S. Army Environmental Health Agency (AEHA) because of lack of funds.

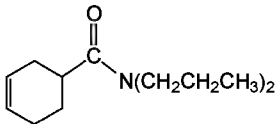
More recently the *cis* and *trans* isomers of the mosquito repellent CIC-4, a mixture of citronella isomers, have been separated by preparative hplc and bioassayed for effectiveness (23). Chiral-phase capillary gas chromatography and mosquito repellent activity of some oxazolidine derivatives of (+)- and (-)-citronellal have been studied to find structure–activity relationships (24). Several 2-alkyl-*n*-acetyloxalidines have been synthesized and tested against mosquitoes, with further efforts using nmr to determine the rotational isomers of the more active *N*-acetyl-2,2-dimethyloxazolidine (25).

Natural Mosquito Repellents. Thousands of exotic and native natural botanicals, plant extracts, and 12,000 synthesized compounds have been collected worldwide and tested for activity during and after World War II in Orlando and Gainesville, Florida by the U.S. Department of Agriculture (USDA). A USDA review of plants of possible insecticidal value lists 1182 species, and is followed by two other compilations of botanicals tested on a large scale by USDA from 1940 to 1964 (9). The review includes tests against 15 species of insects. A further study of synthetic compounds has been published (26). Natural and synthetic repellents have been reviewed in a comparison with synthetic repellents (14), which include common materials such as pyrethrum and Vitamin B₁, as well as many essential oils, including citronella, *Artemesia*, lemon, eucalyptus, mint, and *Clausena*. Cedarwood oil and citronella have been tested against four African species of mosquitoes in Tanzania in competitive tests with DEET, dimethyl phthalate, ethylhexanediol, and permethrin (14). A Chinese lemon eucalyptus oil formulation, Qwenling, received some attention and was found to have effectiveness but only for a short (10–20 min) time compared to a standard DEET treatment, which can last for several hours (27). This is typical for repellent materials described as natural and obtained from plant essential oils that must be reapplied often.

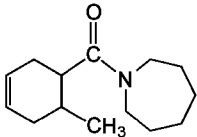
BITING MIDGE REPELLENTS

The genus *Culicoides* is found in fresh water, salt water, and tide water environments in the southeastern United States and Caribbean, where they may be properly called biting midges, or more commonly sandflies, sand fleas, sand gnats, punkies, no-see-ums, or flying teeth. Because of their small size, they can pass easily through ordinary mosquito screens. Commonly used repellents are of varying effectiveness, depending on the species involved (28). In paired field tests, topically applied candidate carboxamide repellents have been compared to DEET at Parris Island, South Carolina, against *C. hollensis*, and on the

Gulf coast of Florida against *C. mississippiensis* (29). Repellency has been determined three ways, ie, by biting rates, length of protection, and coefficient of protection. DEET averaged 61 times greater protection than the untreated check, and all others averaged 130 times greater protection at Parris Island. In Florida, DEET averaged 28 times greater protection than an untreated check, whereas all other materials averaged 190 times greater protection. DEET was considered an effective repellent when applied to exposed skin as a 25 wt % formulation, but four novel alicyclic piperidines (carboxamides) have been found more effective than DEET: *N,N*-dipropylamino-3-cyclohexenecarboxamide [68571-08-4] (6), bp 93°C (60 Pa (0.45 mm Hg)); 2-methyl-1-piperidyl-3-cyclohexenecarboxamide [69462-43-7] (5, R = CH₃), bp 110°C (103 Pa (0.77 mg Hg)); 1-piperidyl-3-cyclohexenecarboxamide [52736-58-0] (5, R = H), bp 108°C (60 Pa (0.45 mm Hg)); and 1-(hexahydro-1*H*-azepinyl)2-methylcyclohexenecarboxamide [67013-96-1] (7), bp 115°C (120 Pa (0.9 mm Hg)). The first two were substantially more effective against *Culicoides* species.



(6)



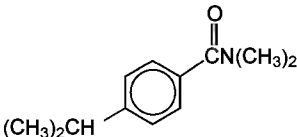
(7)

Among the 117 compounds synthesized, structures that have been found especially active include the 1-piperidyl- (5) and 1-[hexahydro-1*H*-azepinyl] (7) derivatives where the cyclohexenyl ring can either be monounsaturated or have methyl branches such as in the 2-position (30). The use of Avon Skin-So-Soft, a scented mineral oil cosmetic, has been claimed to repel biting midges at Parris Island, South Carolina, and it is reportedly widely used for mosquitoes as well (28). When applied liberally, it apparently traps the midges or fouls their mouthparts and thus inhibits biting. Tests of several commercial mineral oil preparations, both scented and unscented, show interference with biting behavior and therefore a repellent effect. This is because when applied thickly to skin, the small biting midges tend to become thickly coated and drown in the oil (28).

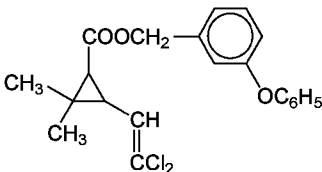
PHLEBOTOMINE SANDFLY REPELLENTS

Because military and civilian personnel were potentially exposed to biting insects during involvement of Allied military in the Persian Gulf and Middle East in the 1980s, U.S. personnel had access to DEET and permethrin clothing treatment through the military supply system. Phlebotomine sandflies are found primarily in relatively underdeveloped tropical and subtropical regions of the world. They are vectors of the various forms of leishmaniasis, bartonellosis, and numerous arboviruses, including the medically important sandfly fever group. However, personal protection against phlebotomine sandflies has had little attention as compared to repellents for mosquitoes. Only a few tests of repellents for Old World phlebotomine flies have been documented, but none in the 1990s.

In tests done in the 1940s, dimethyl phthalate and pyrethrum cream were found to be partially protective on skin against *Phlebotomus papatasi* for 6 h (31). Tests with laboratory-reared *P. papatasi* show that the duration of complete protection (no bites) provided by DEET, *o*-ethoxy-*N,N*-diethylbenzamide, *o*-chloro-*N,N*-diethylbenzamide, or *N*-butyryl-1,2,3,4-tetrahydroquinoline, averages at least 4 h, but perspiration contributes to a high rate of repellent loss (32). Investigations using repellent-treated netting indicate that DEET-treated bednets provide complete protection throughout the night (33). In Panama, phlebotomine sandflies are responsible for the transmission of *Leishmania braziliensis panamensis*, the causative agent of cutaneous leishmaniasis, a human disease among people living in close association with the forests. Cutaneous leishmaniasis and arboviruses affect locals and U.S. military personnel who train in jungle areas. A more recent study shows that 1.6% of U.S. Army personnel who took part in the jungle warfare training course during a three-week period contracted leishmaniasis (34). Personal protection methods have been evaluated against phlebotomine sandflies in the bush in Panama. Arms are treated with 2 mL of 12 wt % max of the following repellents in paired tests with DEET, which was applied to the other forearm in the same manner: hexahydro-1-((2-methylcyclohexyl)carbonyl)-1*H*-azepine [52736-62-6] (7, saturated six-membered ring); 4-isopropyl-*N,N*-dimethylbenzamide [6955-06-2] (8); DEET (1); and *N,N*-dipropylcyclohexanecarboxamide [67013-94-9] (6, saturated ring). Skin applications of the five selected repellents including DEET provide a mean coefficient of protection (CP) of 99.2% against the attack of at least three species of *Lutzomyia*. All of these repellents tested at the highest dosage give good protection from the bites of at least three species of phlebotomine sandflies, two of which are important vectors of leishmaniasis. However, the azepine (7, saturated) gives complete protection and warrants further study (28).



(8)



(9)

DEET-treated net jackets also provide good protection, but an additional application of 10–25% solutions of repellent to the unprotected face is necessary for maximum protection. Clothing treated with permethrin [52645-53-1] (9) does not provide the protection expected against these insects. Because sandfly behavior and resistance to quick knockdown are responsible for the numbers of bites recorded, maximum protection from bites thus requires application of DEET or another suitable repellent to the exposed skin when wearing permethrin-treated clothing (35).

HORSEFLIES, GREENHEADFLIES, DEERFLIES, STABLEFLIES

Field trials of permethrin-treated clothing against highly susceptible tsetse flies have shown good effectiveness (36). There are no published tests exhibiting the effectiveness of clothing treatments against North American species of horseflies and deerflies. Bath oils are commonly used by lifeguards on the Florida Gulf Coast for protection against biting stableflies, and have been used effectively on horses; however, as usual, repellents fail at high fly populations. This repellent effect is difficult to demonstrate in cage tests with stableflies, where biting attacks occur despite treatment with this material (37). It is likely that use of repellent-treated mesh jackets and hats for mosquitoes and sandflies would be most effective against these large biting flies. Table 4 compares the effectiveness of eight repellents to DEET.

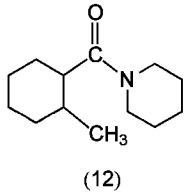
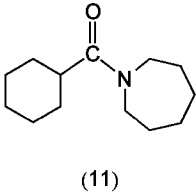
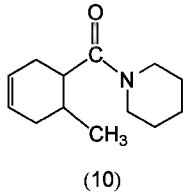


Table 4. Relative Effectiveness^a Compared to DEET^b

Priority	Structure number	Name	CAS Registry Number	<i>Aedes aegypti</i>	<i>Aedes taeniorhynchus</i>	<i>Anopheles quadrimaculatus</i>	<i>Stomoxys calcitrans</i>	<i>Chrysops atlanticus</i>	Blackflies	<i>Culicoides</i> (sandflies)
1	(5, R = H)	1-(3-cyclohexen-1-yl)carboxamide	[52736-58-0]	<	>		>	>	>	>
2	(6, saturated ring)	<i>N,N</i> -dipropylcyclohexanecarboxamide	[67013-94-9]	<	>	<	>	>	<	>
3	(7, saturated ring)	hexahydro-1-((2-methylcyclohexyl)-carbonyl)-1 <i>H</i> -azepine	[52736-62-6]	<	—	<	>	>	=	>
4	(7, no methyl)	1-(3-cyclohexen-1-yl)carboxamide	[52736-59-1]	<	<		>	=	>	
5	(8)	<i>p</i> -isopropyl- <i>N,N</i> -dimethylbenzamide	[6955-06-2]	<					<	>
6	(10)	1-((6-methyl-3-cyclohexen-1-yl)carbonyl)pyrrolidine	[67013-95-0]	<		<		=	=	
7	(11)	1-(cyclohexylcarbonyl)-hexahydro-1 <i>H</i> -azepine	[68571-09-5]	<		<	<			
8	(12)	1-((2-methylcyclohexyl)carbonyl)pyrrolidine	[52736-60-4]		<					

^a >, <, or — indicates statistically significant differences of greater than, less than, or equal to DEET, respectively.

^b Ref. 38.

TICK AND CHIGGER REPELLENTS

Prevention of tick attachment is possible by mechanically preventing access of the ticks to bare skin bordering or beneath the clothing, such as zippered long pants with bloused cuffs tucked into boot or sock tops and a long-sleeved shirt tucked into the pants (2). Repellents are best impregnated into clothing, on wrist skin under the sleeve, on and above the socks, and around the neck on the exposed skin and under the collar. DEET products are available as a 50% liquid and may be mixed with isopropyl alcohol (59 mL DEET and 1 L 2-propanol) to produce sufficient material to impregnate clothing with a 5% solution and, depending on tick density, give 80–98% protection.

Permethrin was developed as a clothing impregnant for military use by several countries worldwide, including the United States. It first became available in 1982 as a clothing treatment under EPA label 24C in an aerosol formulation called Permethrin Tick Repellent. The same formulation is widely available in products registered with approved labeling under such trade names as Duranon Tick Repellent and Coulston's Permethrin Arthropod Repellent. A topical permethrin treatment of clothing has shown good effectiveness against crawling insects when applied as a water-based formulation of 0.5 wt % permethrin (38). It gives extremely effective protection against ticks. Permethrin-treated cloth is practically odorless; a single treatment of ca 20-s spray (ca 0.2 g m²) adsorbs to the outer surface of clothing, does not contact skin, and is long lasting (38). Permethrin as a clothing treatment acts more as a toxicant than as a repellent, and though ticks may crawl on the clothing, the visit is only temporary and usually fatal within a few minutes. The lethal barrier provided has been shown to give 100% protection in tests against ticks in Oklahoma, Kentucky, and Florida (38). The materials named above are

likewise effective repellents against chigger mites (*Trombiculidae*), also called chiggers or redbugs.

BODY AND HEAD LICE

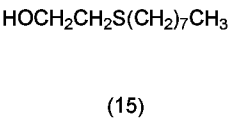
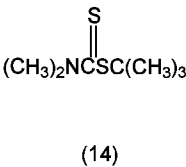
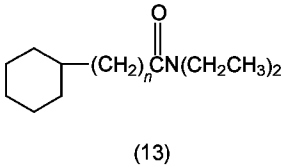
The human body louse, *Pediculus humanus* L., is an important health threat in gyms and schools of the United States and in many areas of the world. There are no licensed vaccines for typhus or relapsing fever. Even if vaccines were available, protection of military personnel would not prevent epidemics in refugee or POW camps under U.S. military control. Lindane was the standard U.S. military pediculicide for over 20 years, used as a delouser during the Korean conflict and on Iraqi troops in early 1991 (22). However, it is believed by some to be obsolete, and is considered a possible oncogen by the U.S. EPA. Also, most populations of body lice are resistant to the organochlorine insecticides, including lindane. The all-purpose clothing impregnant, M-1960, was developed for the U.S. Department of Defense (DOD) and introduced in 1953. It had poor user acceptance on account of its plasticizing properties, disagreeable odor, and irritation to sensitive skin, and is no longer manufactured.

Alternatively, fabric patches treated with permethrin have been evaluated against natural and laboratory strains of human body lice in Peru. Permethrin-treated fabric is toxic to lice on contact and quickly affects feeding behavior, even when washed up to 20 times. Thus permethrin-treated clothing interrupts disease transmission, and offers a passive louse control not previously feasible (39).

Permethrin, under consideration by DOD as a candidate pediculicide for emergency louse control, is marketed as a 1% cream rinse for head louse control. It has been successfully used as a dust formulation against body lice in Egypt. During World War II, studies were done in the United Kingdom, the former Soviet Union, and the United States on the use of various chemicals for impregnating underwear to prevent louse infestations. Pyrethrins have been found effective, but only at high rates of application, and are mostly removed by laundering (39).

COCKROACH REPELLENTS

General information on cockroach control, including repellents and toxicants, is available (40). Transport of goods and materials also provides rapid transport of cockroaches in corrugated cardboard boxes, empty beer and soft-drink bottles, cases in recycling locations, and commercial trucks used for transporting commodities such as bananas, laundry, dry cleaning, and paper bags. Personal automobiles also helped in the rapid dispersal of a newly introduced pest, the flying Asian cockroach, across central Florida in the late 1980s. Repellents may be helpful in preventing transport of cockroaches into uninfested areas. Some logical uses of repellents are on cardboard cartons for food and soft drinks, on beer crates, and in coin-operated vending machines, all of which provide excellent shelter and food for cockroaches (40). Recycling of beer cans and soft-drink containers offers cockroaches another opportunity for shelter and transport, and control is probably difficult. A good repellent can be used either alone or in conjunction with an insecticide as a residual treatment in business establishments or homes. Such effective, long-term repellents can become more useful in the future if the only toxicants available are short-term biodegradable materials. This is especially problematic when retreatment is expensive and rapidly becomes ineffective. Also, the cockroach's opportunistic nature of feeding and shelter-finding permits survival and flourishing when most but not all sites are treated. Similarly, the use of slow-acting toxicants such as borax and boric acid is not effective for long unless insects can be confined to dry, treated surfaces. This tends to describe a laboratory environment and is not applicable to the real world in which cockroaches may quickly leave an effectively treated area and fully recover from the sublethal effects. Many repellents are found among amides, sulfonamides, cyanoacetic acids, and carboxamides, but two good ones are *N,N*-diethylcyclohexanecetamide (13, *n* = 1) and *N,N*-diethylcyclohexanepropanamide (13, *n* = 2), both better than fencholic acid when tested against the common North American cockroaches, *Blattella germanica* and *Periplaneta americana* (41).



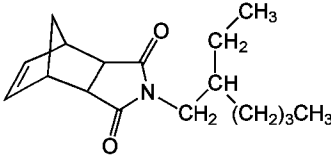
Dibutyl succinate or Tabutrex [141-03-7]; R-11; R-55 [23885-27-0] (14); and R-874 [3547-33-9] (15) have been available for industry as cockroach repellents since the 1960s. Tabutrex (Glenn Chemical Company) is formulated as an emulsion concentrate (20%) and an oil spray (2%). The oral LD₅₀ (rat) is 8000 mg/kg. Treated surfaces remain 100% repellent to *B. germanica* for three weeks. In laboratory tests, cockroaches are repelled from wooden beverage crates for 15 weeks (42). Hexahydrodibenzofurancarboxaldehyde–butadienefurfural copolymer, MGK R-11 (3) (Phillips Petroleum Company) is a pale yellow liquid having a fruity odor, miscible with many organic solvents, and compatible with most insecticides. A typical formulation contains 0.075% pyrethrins, 0.15% piperonyl butoxide, and 1% R-11. For treating the inside of cartons, R-11 is applied as a 1% emulsion incorporating 2% of the synergist MGK 264. On beer cartons, R-11 gives >80% repellency for two months, reducing to 60% at six months. MGK R-11 has been used in pet sprays and in repellents for personal use. Of all the materials evaluated for odor, this repellent is the most pleasant (43). The acute oral LD₅₀ (rat) is 2500 mg/kg; the dermal LD₅₀ is >2000 mg/kg.

t-Butyl *N,N*-dimethyldithiocarbamate (14) or MGK R-55 (McLaughlin Gormley King Company) is a rodent and insect repellent. It repels *B. germanica* from treated cartons for 90 d (at 2%) and for 63 d (at 1%). It is more odorous and toxic than MGK R-11 and MGK R-874. However, 2-hydroxyethyl *n*-octyl sulfide (15) or MGK R-874 (Phillips Petroleum Company), the only commercially available repellent, is a light amber liquid having a mild mercaptan-like odor, slightly soluble in water but miscible with most organic solvents (40). The label indicates that it may be used near food (40). It is used with MGK 264, a pyrethrins synergist. Formulations commercially available are an EC diluted with water and applied at 1–5% by automatic spraying equipment and an oil solution used at one gram of active material per square meter. R-874 tested against German cockroaches is marginally more effective than R-55 and lasts twice as long as R-11. Toxicity is low; the acute oral LD₅₀ (rat) is 8530 mg/kg; dermal LD₅₀ is 13,590 mg/kg.

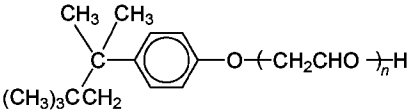
A listing of compounds evaluated in the laboratory as cockroach repellents summarizes 872 synthetic compounds out of 901 bioassayed from 1953 to 1974 (43). Fencholic acid [512-77-6] (3-isopropyl-1-methylcyclopentanecarboxylic acid) has been used as a standard repellent in tests conducted by placing 20 cockroaches in a glass crystallizing dish without food and water and offering them a choice of two cardboard shelters, one of which was treated with 1 or 2 mL of a 1% solution of the candidate in acetone. Counts were made daily for seven days.

Another problem lies in the overlap of repellent–toxicant definition, in that many toxicants are known to have repellent effects (43). Pyrethrins are

often used on ships to flush cockroaches from harborages during a treatment with another, less activating toxicant. In a survey of the components, eg, toxicants, synergists, solvents, flushing agents, and emulsifiers, making up commercially available formulations of insecticides for cockroach control in the United States, 121 different materials were examined (44). Tests show that pyrethrins which have been considered repellents for some years, MGK 264 [113-48-4] (16) and the emulsifier Triton X100 [9002-93-1] (17), are noticeably repellent to both German and American cockroaches (44).

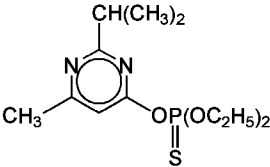


(16)

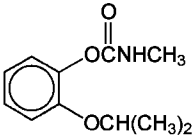


(17)

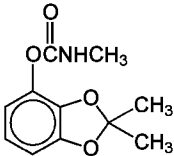
The list of repellent materials also includes a number of surfactants (wetting agents) and deodorants, but in no case are solvents implicated (45). In laboratory studies for repellency, some formulations containing 0.5% organophosphates did not function as repellents, but diazinon [333-41-5] (18) (0.5%), propoxur [114-26-1] (19) (1%), synergized pyrethrins (1%), some synthetic pyrethroids, and bendiocarb [22781-23-3] (20) (1%) were repellent for a week or more (46). In an extensive testing program of many insecticides, avoidance of treated surfaces has been observed more frequently with diazinon than with any of the other materials (47). Diazinon (18) is commonly used in Florida for household treatments, although chlorpyrifos, permethrin, cypermethrin, and hydroprene are widely used for cockroach control.



(18)



(19)

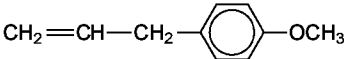


(20)

Sixty-two novel experimental carboxamides of 1,2,3,6-tetrahydropyridine have been tested as repellents of German cockroaches, and five provided 100% repellency for 17 d in a stringent test (48).

OTHER INSECTS

Bark beetle management in European forests has been successful using combinations of sex pheromones and tree volatiles. Repellents that were tested in Louisiana to deter attacks of the southern pine beetle afforded protection of high value loblolly pines by using the host tree compound 4-allylanisole [140-67-0] (49). The aggregation inhibitor 4-allylanisole (21) eliminated tree deaths for the length of the 30-d test by placing nine vials with wicks containing 20 g each of repellent vertically on the lower trunk of each tree being protected, using the tree as a flagpole. A patent has been issued on this technology (49).



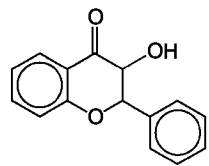
(21)

Bird Repellents

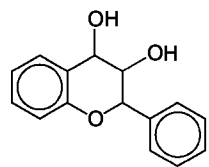
Blackbirds, starlings, and sparrows are North American birds that cause serious damage to growing crops, costing at least \$40 million yr. Nonchemical techniques using repelling devices such as propane cannons, shiny Mylar ribbons, scarecrows, metallic pinwheels, and recorded distress calls give temporary results, but when the birds become accustomed to the devices, the effect is generally lost (50). However, when reflective tapes were stretched at close

intervals over entire fields of a high value sweet corn crop, losses of corn ears to blackbirds were one-sixth to one-third of losses in untaped fields; goldfinches and deer were not deterred (50).

Millet is a grain-yielding sorghum, a vital staple food crop occupying 44×10^6 ha (10.9×10^7 acres) in the Third World, including India, southern Asia, Latin America, the Sahelian zone of Africa, the Near East, and the Middle East. The main bird pest in Africa is *Quelea quelea*, a weaver finch. In many of these areas where control measures are necessary for the preservation of the crop (51), chemical repellents are expensive and difficult to obtain, require special application equipment, and therefore in some situations are an unlikely consideration. For these areas, it seems practical to breed the ability to resist bird depredation into the physical characteristics of the plants (52) or the genetic composition of the plants, and much effort has been so directed since 1960 (53). High content of tannins is the characteristic most often associated with bird resistance in sorghum because these polyphenolics (tannins) produce astringency and thus repellency. Unfortunately, the palatability, digestibility, and nutritional quality of foods may also be reduced in tannin-loaded food products. Hydrolyzable tannins are present in small quantities in sorghum, and condensed tannins are responsible for coagulation of proteins of the saliva and mucous membranes, resulting in the astringent taste response. Polyphenolic condensed tannins or proanthocyanidins are a series of complex condensed 4-ketoflavan-3-ol [577-85-5] (22) and flavan-3,4-diol [5023-02-9] (23) molecules of 500–3000 mol wt (54). The subject of polyphenolic tannins has been reviewed (55); however, application of natural tannins onto crops failed to show efficacy.



(22)

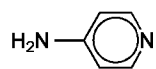


(23)

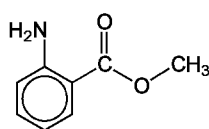
Some bird repellents are composed of viscous, sticky materials that birds dislike having on their feet (17). These compounds, eg, Tanglefoot, Roost-No-More, and TackTrap, are often based on incompletely polymerized isobutylene and thinned with aromatic solvents. They should be formulated to have the proper blend of tackiness and viscosity for the weather, method of application, and pest species. They are applied to leave sticky residues on perching locations in buildings and roosts in trees. Because these materials do not have an obnoxious odor, the birds must land on and learn its location in order to avoid it, as there are no long-range cues in the treatment itself for conditioning.

Intoxicating chemicals are those that are not necessarily lethal (see PESTICIDES) but operate as primary repellents or secondary repellents, eg, emetics causing sickness or distress. Primary bird repellents are those whose mode of action is having a bad taste; immediate rejection of food is the desired result. However, they are effective only if other foods are available; they are not effective in times of food shortages, because large flocks of migrating birds would be forced to feed or starve. Bird repellents have been discussed in reviews (51,56).

Avitrol [504-24-5] (4-aminopyridine) (24), mp 155–158°C, bp 273°C, has repellent–toxicant properties for birds and is classed as a severe poison and irritant. This secondary bird repellent can be used as a broadcast bait, causing uncoordinated flight and distress calls and escape responses in nearby birds (57). A reevaluation shows lack of effectiveness of 1% baits but better control of blackbirds with 3% baits (58). Suspected contamination of drinking water with 4-aminopyridine has been reported in toxicosis of Brahman cattle and horses (59).

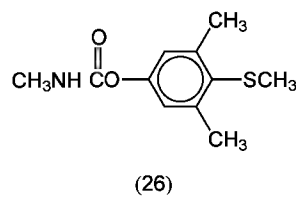


(24)



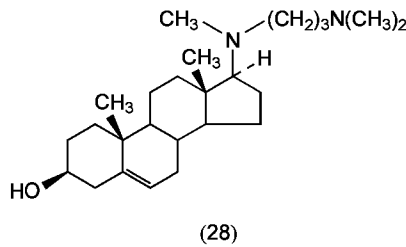
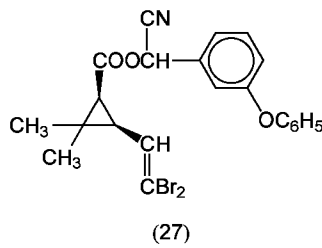
(25)

Methyl anthranilate [140-20-3] (25), the grape flavoring used in food products, has been shown effective as a waterfowl repellent when applied at 90–169 kg/ha (8–15 lb/acre). Research has shown statistically significant reduction of activity compared to untreated water with 0.06–0.5% solutions applied onto shallow standing water next to airport runways (60). In more recent efforts, a free-flowing powder formulation was added to 1-m dia children's wading pools (at 0.075% w/w) and showed significant reduction of activity (94–96% less activity) against free-ranging gulls for 4–11 days, compared to untreated water. Overall gull activity has been reduced even when all water was treated (61). A surfactant-containing formulation was tested against mallard ducks in 1-m dia wading pools at 0.02% AI, and both pool entry and bill dipping were measured and found to be significantly reduced (61). These materials demonstrated repellency at concentrations of 0.038% vol/vol, which are 10–60 times lower than concentrations needed to repel red-winged blackbirds and European starlings from solid livestock feed (61). Also, data collected support evidence of long-lasting effects and suggest learned avoidance of anthranilate compounds by birds (62), a further indication that these compounds may be useful in reducing damage to newly planted rice fields and to reduce losses at fish hatcheries.



Methiocarb [2032-65-7] (3,5-dimethyl-4-(methylthio) phenol methylcarbamate) (26) is classed as an insecticide and acaricide and is used as a slug and snail bait, but is no longer registered for use as a bird repellent in the United States. Its uses on field and horticultural crops for bird repellency as an emetic have been reviewed (63). It was found to reduce bird damage in treatments of sweet corn (64). Methiocarb has been applied to wine grapes in Ohio, California, and Oregon (65), and to blueberries in New Zealand (66). Residues in wine (qv), as well as its effect on the composition and flavor of the bottled wine, were reported (67). Its efficacy in ripening sorghum in Canada and Senegal were also reported (51,68), as were its residues and its sulfoxide and sulfone metabolites during efficacy studies against starlings in cherry orchards (69). Sorghum hybrids were treated with methiocarb, and grain yield and predation were studied (70). The conditioning response acquired is effective against red-winged blackbirds and persists in the laboratory up to 16 weeks (71). More recent studies to answer EPA queries show lack of methiocarb toxicity to birds and mammals in the laboratory and during field studies in fruit and sweet corn using labeled treatment levels. Based on estimates from 26 studies, treated plantings average 15% loss of fruit to birds compared to 36% for nearby orchards; it has been concluded that methiocarb has efficacy in repelling birds from fruit crops when applied at 1.7 kg/ha, a level that does not adversely affect birds (71). Calcium carbonate has been added to methiocarb in an effort to increase its effectiveness as a visual cue, but failed to enhance bird repellency in ripening sorghum (72).

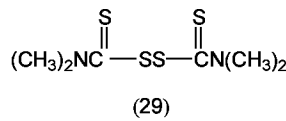
Anthrahydroquinones have been patented in Japan as bird repellents (73), and anthraquinone [84-65-1] (qv) is used widely in Europe as a spray to protect growing crops and as a wood dressing. The synthetic pyrethroid deltamethrin [52918-63-5] (27) was evaluated (74), as were other materials, including bendiocarb (20) (75) and 20,25-diazocholesterol dihydrochloride [1249-84-9] (Omitrol) (28), a steroid that inhibits embryo development when adsorbed or ingested as a seed treatment of bait corn (55,76).



Mammalian Pests

The concept of employing a nonlethal repellent to control wildlife depredation on crops arose early in agricultural history and has been pursued vigorously ever since. Although the continued interest in repellents may reflect public opinion about the impact on endangered or protected species, feeding inhibitors and modern lethal treatments remain practical solutions. A food repellent has been defined as "a compound or combination of compounds that, when added to a food source, acts through the taste system to produce a marked decrease in the utilization of that food by the target species" (50). The action can be primary, where the animal reacts to the taste of the repellent alone, or secondary, where the animal uses the taste of the repellent as a cue to later adverse effects. A useful repellent is meant to stop a hungry animal from feeding on a readily accessible, abundant, and palatable food, forcing the pest animal to leave the area or make a change in food habits, both unlikely choices. The feeding activity of deer has become an increasingly important problem in the U.S. Pacific Northwest, where black-tailed deer and Roosevelt elk browse Douglas fir seedlings. Nonlethal repellents to protect crops from vertebrate pests, together with some considerations for their use and development, have been reviewed (50,77).

Evaluations have been conducted using deer, a multiple-choice preference-testing apparatus, and tetramethylthiuram disulfide [137-26-8] (TMTD) (29) or the fungicide thiram as a standard repellent for competitive tests with repellent-treated food (78).



A fermented-egg product (FEP), patented as an attractive bait for synanthropic flies, has been shown to be attractive to coyotes and repellent to deer (79). Its components are variable, with relative concentrations of 77% fatty acids, 13% bases, and 10% (primarily) neutrals composed of at least 54 volatiles such as ethyl esters, dimethyl disulfide, and 2-mercaptoethanol. Synthetic formulations have been evaluated to find a replacement for a patented fermented-egg protein product that attracts coyotes and repels deer. Ten aliphatic acids (C-2 to C-8), four amines (pentyl, hexyl, heptyl, and trimethyl), dimethyl disulfide, 2-mercaptoethanol, and 54 more volatiles (C-1 to C-5 esters of C-1 to C-8 acids) have been tested as synthetic fermented egg (SFE) (80) in approximately the same proportions that are present in FEP. Weathering was a problem that caused decreased efficacy, which suggests trials of controlled-release formulations. Fourteen repellents have been examined against white-tail deer in Pennsylvania in choice tests when treated onto shelled corn (81).

Hinder or Repel, registered under the Federal Insecticide, Fungicide, and Rodenticide Act (FIFRA) Section 24C, a state registration for special local need only, repels deer and rabbits from fruit trees, vines, vegetables, field crops, forage and grain crops, ornamentals, nursery stock, and noncrop areas. It is best applied before damage occurs as an aqueous spray or by painting and is claimed to last 3–8 weeks. Hinder contains 15% ammonium soaps of higher fatty acids (1.5% ammonia and 13% mixed rosin and fatty acids) and 85% inert ingredients (81). The material is sold in the western United States as Hinder and in the eastern United States as Repel or Sticker-spreader 268. Chaperone is the only material as of this writing (1996) approved by EPA in Florida as a repellent for deer, mice, and rabbits. About 10 materials have been registered in Georgia in 1982, usually containing 4–22% thiram (29) (82).

Although no consistently effective chemical repellent has been developed for vertebrate pests, some promising materials have been tested as repellents that are based on predator avoidance, specifically compounds from the secretions of predators. In 1995, synthetic sulfur compounds (two thietanes, a thiolane, and a substituted methyl sulfide, which were originally identified from the anal glands of the stoat, ferret, and red fox) suppressed browsing by the introduced Australian brush-tail opossum in New Zealand about as well as FEP (83). Suggestions were made that these compounds can be made more effective by the use of bitter compounds in a cocktail.

Area repellents are materials that are intended to keep animals away from a broad area. They include predator scent such as lion or tiger manure, blood meal, tankage such as putrefied slaughterhouse waste, bone tar oil, rags soaked in kerosene or creosote, and human hair (84). Although few controlled tests have been run on these materials in the past, more recent investigations of predator odors have shown promise (85).

Health and Safety Factors

Toxicology. Toxicological testing has been carried out on many of the older, widely used materials, all of which require re-registration with the EPA (86). This accounts for the disappearance from the U.S. market of 2-ethylhexane-1,3-diol [94-96-2]. Few of the newer compounds have been submitted for extensive toxicological testing because of cost, problems of registration (87), and a necessity to be competitive in the marketplace with every new product. As a result of EPA regulations, many of the materials submitted as cloth repellents since 1970 have been tested at the USDA Agriculture Research Service, Medical and Veterinary Entomology Research Laboratory in Gainesville, Florida. Effective compounds, after further testing, are then submitted to the U.S. Army Environmental Health Agency for extensive toxicological testing. Compounds are tested as repellents on human skin only after passing the four standard toxicological tests: rabbit eye irritation, rabbit skin dermal, rat inhalation, and rat acute ingestion. All of these, plus EPA regulations in the United States classifying repellents as pesticides, have drastically reduced the number of candidate chemicals submitted to the USDA laboratory in Gainesville (2) for general screening since about 1975, and virtually eliminated chemicals submitted as candidate repellents. As a result, this function of the USDA may be eliminated. Some materials of either private or public origin continue to be tested in the 1990s under a Cooperative Research and Development Agreement (CRADA) system (88). Canada Health and Welfare and Occupational Health have tested DEET for skin penetration on the forehead of monkeys and claimed that it was toxic (89). As a result, most products having high concentrations of active ingredients are either canceled in Canada, or require warning labels against application to bare skin.

Hazard Assessment of Chemical Repellents. Labels for repellent products sold in the United States are recommended for purposes of efficacy and safety of use. Newer products containing DEET may contain less active ingredient but feature a cosmetic that makes the compound less objectionable on the skin and more acceptable to use (2). Even though such a treatment may last for less time, it may help decrease exposure and potential adverse effects, especially on children and or adults with sensitive skin. NIOSH has recommended for National Park Service employees of the Everglades National Park in Florida that DEET use should not exceed the amount absolutely necessary for repellency (90). Serious adverse reactions are rare to DEET (91) unless used to drastic excess. Since 1954, six female children under the age of eight have been reported with toxic encephalopathy associated with use of products containing DEET. Generally the children had been excessively overtreated from three days to three months, thus resulting in three deaths; however, the causes of death have not been resolved.

The dermal adsorption of DEET in humans has been studied in the Netherlands by application of [¹⁴C] DEET as undiluted technical material or as 15° solutions in alcohol. Labeled material was recovered from the skin, and absorption of DEET was indicated by the appearance of label in urine after two hours of skin exposure. About 5–8° of the applied treatments was recovered as metabolites from urine, and excretion of metabolites in the urine came to an end four hours after exposure ended. DEET did not accumulate in the skin, and only a small (less than 0.08°) amount ended up in feces. Curiously, less has been absorbed through skin from 100° DEET application (3–8°, mean of 5.6°) than from 15° alcohol application (4–14°, mean of 8.4°). These results have been described as consistent with previous absorption metabolism studies using guinea pigs, rats, and hairless dogs. Other publications on DEET toxicology have been cited (92).

Dog repellents available commercially in the 1990s have been generally unsuccessful in laboratory tests. For example, lithium chloride treatments were usually rejected immediately with no ingestion, and bone oil treatments that contained up to 0.1° of the active ingredient were still consumed (93). Oleoresin capsicum [8023-77-6], the essence of red pepper, did have an extended effect on coyotes, even though the deer repellents mentioned above were attractive to coyotes (93). Although a capsicum-base aerosol repellent has been described as potentially harmful (94), pepper spray is commercially available in the United States to repel humans, as is Mace.

Numerous articles in the popular press have stated that heavy consumption of vitamin B₁ (thiamine) can stop attacks of biting and stinging insects on the thiamine-loaded human (see VITAMINS, THIAMINE). This was investigated during World War II, in post-war tests (95), and as recently as 1992 at Gainesville (22). There is no scientific evidence that thiamine has any effect whatsoever on the attraction of *A. aegypti* to humans in olfactometer tests, whether taken internally to excess or applied externally, during scientific tests in 1944, 1952, 1969, and 1973 (2). The same results have been noted for gadic by the U.S. Food and Drug Administration, which concluded that, because of the lack of adequate data to establish the effectiveness of this or any other ingredient for over-the-counter (OTC) internal use as an insect repellent, labeling claims for OTC orally administered insect repellent drug products are either false, misleading, or unsupported by scientific data (96).

Mechanical Noisemakers

Claims of effects of repelling or disrupting ultrasonic devices on selected rodent species (97) have been extended by some producers of such devices to include repelling of cockroaches, mosquitoes, fleas, and other insects. There is replicated scientific evidence that shows no effect of several sonic and ultrasonic frequencies (1,000–60,000 Hz) on German cockroaches in choice boxes, because the cockroaches were neither killed nor repelled (98). No effect was seen on fleas or cockroaches (99). Experiments with human arms in olfactometers showed no effect on the attraction of *A. aegypti* when sonic devices were used. Mosquito attraction was statistically the same whether or not any of several makes of small portable sonic devices (600–1000 Hz) reputed to repel mosquitoes were activated (100), regardless of the claims for the production of wavelengths of sound produced by male mosquitoes (98,101). Warnings were sent in the spring of 1993 to some distributors of ultrasonic pest-control devices, which noted that "statements that pertain to the efficacy of the product have not been substantiated and when used in connection with the product could be in violation of the FIFRA" (27).

Extension of Repellent Effectiveness

Attempts to extend repellent effectiveness involve chemical bonding of the repellent molecule to dermophilic compounds that then bind to the skin. Compounds containing 1,3-dihydroxyacetone and pendent repellent molecules were investigated until 1972 (102), as were amino acid analogues of 2-ethyl-1,3-hexanediol, but results were not outstanding (103). Effective cosmetics formulation technology is available in the 1990s to extend the effective length of DEET on skin (2). These materials use extenders and odor-masking agents to make the use of DEET more pleasant.

BIBLIOGRAPHY

"Repellents" in *ECT* 3rd ed., Suppl. Vol., pp. 786–805, by D. A. Carlson, University of Florida.

1. C. F. Curtis and co-workers, *Med. Vet. Entomol.* **1**, 109 (1987).
2. C. E. Schreck, in P. S. Auerbach, ed., *Wilderness Medicine: Management of Wilderness and Environmental Emergencies*, Mosby Co., St. Louis, Mo., 1995.
3. C. E. Schreck, in J. Adams, ed., *Insect Potpourri: Adventure in Entomology*, Sandhill Press, Inc., Gainesville, Fla., 1992, p. 79.
4. *World Health*, 10 (Apr. 1982); E. A. Smith, *Mosq. News*, **42**, 510 (1982).
5. *Biology and Control of Aedes aegypti*, *Vector Topics No. 4 and Dengue Surveillance Survey No. 9*, U.S. Public Health Service, Centers for Disease Control and Prevention, Atlanta, Ga., 1979 and 1983.
6. C. E. Schreck, *Ann. Rev. Entomol.* **22**, 101 (1977).

7. F. E. Kellog and co-workers, *Can. Entomol.* **100**, 763 (1968).

8. C. E. Schreck and co-workers, *Mosq. News*, **37**, 455 (1977).

9. W. V. King, *Chemicals Evaluated as Insecticides and Repellents at Orlando, Fla.*, Ag. Handbook 69, USDA, Washington, D.C., 1954.

10. C. E. Schreck and co-workers, *Soap Cosmet. Chem. Special.*, 36 (Sept. 1982).

11. C. E. Schreck and co-workers, *Am. J. Trop. Med. Hyg.* **31**, 1046 (1982).

12. C. E. Schreck, *J. Am. Mosq. Control Assoc.* **7**, 533 (1991).

13. Long Road Travel Supplies, Berkeley, Calif., (800) 359-6040; Epco Design, Juneau, Ark., (907) 586-1622.

14. C. F. Curtis and co-workers, in C. F. Curtis, ed., *Appropriate Technology in Vector Control*, CRC Press, Boca Raton, Fla., 1989, p. 75.

15. M. I. Hossaine and C. F. Curtis, *Med. Vet. Entomol.* **3**, 367 (1989); C. E. Schreck and L. S. Self, World Health Organization, *Vector Biological Control*, 85.914, 1-6 (1985).

16. C. E. Schreck and D. L. Kline, *J. Am. Mosq. Control Assoc.* **5**, 91 (1989).

17. *Pest Control, Retail Producers Guide*, 20 (Mar. 1983).

18. C. E. Schreck and T. P. McGovern, *J. Am. Mosq. Control Assoc.* **5**, 247 (1989).

19. M. D. Buescher and co-workers, *Mosq. News* **42**, 428 (1982).

20. **U.S. Pat. 4,291,041** (Sep. 22, 1981); **U.S. Pat. 4,356,180** (Oct. 26, 1982); and **U.S. Pat. 4,298,612** (Nov. 3, 1981), T. P. McGovern and C. E. Schreck (to USDA).

21. C. E. Schreck and D. L. Kline, *J. Am. Mosq. Control Assoc.* **5**, 77 (1989).

22. C. E. Schreck, personal communication, Gainesville, Fla., June 2, 1995.

23. J. D. Warthen and co-workers, *J. Chromatogr. Sci.* **590**, 133 (1992).

24. W. G. Taylor and C. E. Schreck, *J. Pharmaceut. Sci.* **74**, 534 (1985).

25. W. G. Taylor and C. E. Schreck, *Pesticide Sci.* **33**, 1 (1991); W. G. Taylor and co-workers, *Can. J. Chem.* **70**, 165 (1992).

26. N. E. McIndoo, ed., "Plants of Possible Insecticidal Value," USDA, Washington, D.C., 1945; "Materials Evaluated as Insecticides, Repellents and Chemosterilants, Orlando and Gainesville, Fla., 1952-1964," USDA, Washington, D.C., 1967.

27. C. E. Schreck and B. A. Leonhardt, *J. Am. Mosq. Control Assoc.* **7**, 433 (1991).

28. C. E. Schreck and D. L. Kline, *Mosq. News*, **41**, 7 (1981).

29. C. E. Schreck and co-workers, *J. Med. Entomol.* **16**, 524 (1979).

30. T. P. McGovern and C. E. Schreck, *Mosq. News*, **40**, 394 (1980); **U.S. Pat. 4,530,935** (July 25, 1985), T. P. McGovern and C. E. Schreck.

31. A. B. Sabin and co-workers, *J. Am. Med. Assoc.* **125**, 693 (1944).

32. M. L. Schmidt and J. R. Schmidt, *J. Med. Entomol.* **6**, 79 (1969).

33. V. M. Safyanova, *Med. Parazitol. Parazit. Bolezni*, **35**, 549 (1963).

34. E. T. Takafugi and co-workers, *Am. J. Trop. Med. Hyg.* **29**, 516 (1980).

35. R. H. Grothaus and co-workers, *Mosq. News*, **36**, 11 (1976).

36. L. L. Sholdt and co-workers, *Med. Vet. Entomol.* **3**, 153 (1989).

37. J. Hogsette, personal communication, USDA, Gainesville, Fla., Nov. 1, 1995.

38. C. E. Schreck and co-workers, *J. Econ. Entomol.* **75**, 1059 (1982); C. E. Schreck and co-workers, *J. Med. Entomol.* **19**, 143 (1982).

39. L. L. Sholdt and co-workers, *Military Med.* **154**, 90 (1989).

40. P. B. Cornwell, *The Cockroach*, Vol. II, Associated Business Programmes, Ltd., London, 1976, pp. 157-190; L. D. Goodhue and G. L. Tissol, *J. Econ. Entomol.* **45**, 133 (1952); P. G. Koehler and co-workers, in K. Storey, ed., *Handbook of Pest Control*, 7th ed., p. 100.

41. B. E. Hagenbuch and co-workers, *J. Econ. Entomol.* **80**, 1022 (1987); **U.S. Pat. 4,621,143** (Nov. 4, 1986), T. P. McGovern and G. C. Burden (to USDA).

42. *Pest Control*, **25**, 22 (1957).

43. *Laboratory Evaluations of Compounds as Repellents to Cockroaches, 1953-1974*, Production Research Report No. 64, Agricultural Research Service, USDA, Washington, D.C., Oct. 1976.

44. B. J. Smittle and co-workers, *Pest Control*, **36**, 9 (1968).

45. NPCA Tech. Release No. 15-69, National Pest Control Association, Vienna, Va., 1969.

46. G. S. Burden, *Pest Control*, **43**, 16 (1975).

47. J. M. Grayson, *Pest Control*, **44**, 30 (1976).

48. T. P. McGovern and G. S. Burden, *J. Med. Entomol.* **22**, 381 (1985).

49. J. L. Hayes and co-workers, *J. Chem. Ecol.* **20**, 1595 (1994); **U.S. Pat. 5,403,836** (Apr. 4, 1995) (to USDA).

50. J. G. Rogers, Jr., in R. W. Bullard, ed., *Flavor Chemistry of Animal Foods*, ACS Symposium Series No. 67, Washington, D.C., 1978, p. 150.

51. R. L. Bruggers, in *Quelea quelea: Africa's Bird Pest*, R. L. Bruggers and C. C. H. Elliot, eds., Oxford Press, U.K., 1989, p. 262.

52. R. A. Dolbeer and co-workers, *Crop Protection*, **14**, 39 (1995).

53. R. W. Bullard and B. Gebrekidan, in Ref. 51, p. 281.

54. R. W. Bullard and co-workers, *J. Agric. Food Chem.* **28**, 1006 (1980).

55. R. W. Bullard and D. J. Elias, *Proc. Inst. Food Technol.* **43** (June 1979).

56. E. N. Wright, ed., *Bird Problems in Agriculture*, British Crop Protection Council 23, BCPC Publications, Croydon, U.K., 1980, p. 164.

57. J. F. Besser, *Proceedings of the 7th Vertebrate Pest Control Conference*, Monterey, Calif., 1976, p. 11.

58. P. P. Woronecki and co-workers, *J. Wildlife Manage.* **43**, 184 (1979).

59. S. S. Nicholson and C. J. Prejean, *J. Am. Vet. Med. Assoc.* **173**, 1277 (1981); G. A. Van Gelder, in P. W. Pratt, ed., *Equine Medicine and Surgery*, 3rd ed., American Veterinary Publications, Santa Barbara, Calif., 1982, p. 197.

60. R. A. Dolbeer, *USDA-APHIS Denver Wildlife Center Animal Repellents Report*, U.S. Armed Forces Pest Management Board, Washington, D.C., 1990, 1996.

61. J. L. Belant and co-workers, *Crop Protect.* **14**, 171 (1995).

62. J. F. Gelahn and co-workers, *Wild. Soc. Bull.* **17**, 313 (1989).

63. F. T. Crase and R. W. Dehaven, in Ref. 57, p. 46.

64. P. P. Woronecki and co-workers, *J. Wildlife Manage.* **35**, 693 (1981).

65. R. L. Hothem and co-workers, *Am. J. Enol. Vitic.* **32**, 150 (1981); *Proc. Bird Cont. Semin.* **8**, 59 (1982).

66. *Proc. N. Z. Weed Pest Control Conf.* **33**, 125 (1980).

67. A. C. Noble, *Am. J. Enol. Vitic.* **31**, 98 (1980).

68. R. R. Duncan, *Can. J. Plant Sci.* **60**, 1129 (1980); G. Gras and co-workers, *Bull. Environ. Contam. Toxicol.* **26**, 393 (1981).

69. *Phytoparasitica*, **8**, 95 (1979).

70. *Argon. J.* **73**, 290 (1981).

71. R. A. Dolbeer and co-workers, *Pestic. Sci.* **40**, 147 (1994).

72. R. A. Dolbeer and co-workers, in R. L. Doty and D. Muller-Schwarze, eds. *Chemical Signals in Vertebrates*, Plenum Press, New York, 1992, p. 323.

73. Jpn. Kokai Tokyo Koho 8183408 (July 8, 1981).

74. *Poult. Sci.* **60**, 1149 (1981).

75. *Res. Discl.* **211**, 420 (1981).

76. R. W. Bullard, in T. E. Acree and D. M. Soderlund, eds., *Semiochemistry: Flavors and Pheromones*, W. de Gruyter & Co., New York, 1985, p. 65.

77. D. Muller-Schwartz, in D. W. McDonald, D. Muller-Schwartz, and S. E. Natynzuk, eds., *Chemical Signals in Vertebrates*, Oxford Press, U.K., 1990, p. 585; R. L. Bruggers and co-workers, *Wild. Soc. Bull.* **14**, 161 (1986); R. A. Dolbeer, *Wild. Soc. Bull.* **14**, 418 (1986).

78. D. L. Campbell and R. W. Bullard, *Proceedings of the 5th Vertebrate Pest Conference*, Fresno, Calif., 1972.

79. **U.S. Pat. 3,846,557** (Nov. 5, 1974), M. S. Mulla and Y.-S. Hwang (to 3M Co.).

80. R. W. Bullard and co-workers, *J. Agric. Food Chem.* **26**, 155 (1978).

81. W. Palmer, *Deer-Away Technical Report*, International Reforestation Suppliers, Eugene, Oreg., 1980.

82. J. Jackson, *Deer and Rabbit Repellents*, Dept. of Forest Resources, University of Georgia, Athens, Ga., 1982.

83. *Extension Publication 18*, No. 11, Dept. of Natural Resources, NYSC Agriculture and Life Sciences, Cornell University, Ithaca, N.Y., 1980; *Supplement No. 120*, Extension Wildlife and Sea Grant, University of California, Davis, Calif., Oct. 1979; *Extension Information Bull. No. 146*, Cornell University, Ithaca, N.Y., 1978.

84. A. D. Woolhouse and D. R. Morgan, *J. Chem. Ecol.* **21**, 1571 (1995).

85. R. A. Bruggers, personal communication, Denver, Colo., Jan. 15, 1996.

86. M. L. Leng, in G. J. Marco, R. M. Hollingsworth, and J. R. Plimmer, eds., *Regulation of Agrochemicals: A Driving Force in their Evolution*, ACS Non-Symposium Series, American Chemical Society, Washington, D.C., 1991, p. 26.

87. EPA: *N,N*-Diethyl-m-toluamide (DEET), Pesticide Registration Standard, U.S. EPA, Washington, D.C., 1980.

88. D. R. Zimmer, personal communication, USDA, ARS, Athens, Ga., Jan. 15, 1996.

89. R. P. Moody and co-workers, *J. Toxicol. Environ. Health* **26**, 137 (1989).

90. R. McConnell and co-workers, HETA 83-085-1757, U.S. Dept. Health and Human Services, CDC, Cincinnati, Ohio, 1986.

91. E. H. Roland and co-workers, *Can. Med. Assoc. J.* **132**, 155 (1985).

92. S. Selim and co-workers, *Fund. Appl. Toxicol.* **25**, 95 (1995).

93. Personal communication, R. Teranishi, USDA Western Regional Laboratory, Albany, Colo., 1983.

94. *Vet. Human Toxicol.* **22**, 18 (1980).

95. H. J. Maasch, *Tropenmed. Parasitol.* **4**, 119 (1973).

96. *Federal Register* 48:26987, Part III, Dept. Health and Human Services, June 10, 1983.

97. A. V. Scalingi, *Pest Control*, **48**, 26 (1980).

98. C. E. Schreck, J. C. Webb, and G. S. Burden, *J. Environ. Sci. Health A*, **19**, 521 (1984).

99. P. G. Koehler and co-workers, *J. Econ. Entomol.* **79**, 1027 (1986).

100. W. A. Foster and K. R. Lutes, *J. Amer. Mosq. Control. Assoc.* **1**, 199 (1985).

101. D. J. Lewis and co-workers, *Can. Entomol.* **114**, 699 (1982).

102. R. P. Quintana and co-workers, *J. Econ. Entomol.* **65**, 66 (1972).

103. R. P. Quintana and co-workers, *J. Med. Chem.* **15**, 1073 (1972).

D. A. Carlson
University of Florida

REPROGRAPHY.

See ELECTROPHOTOGRAPHY; THERMOGRAPHY.

RESEARCH MANAGEMENT.

See RESEARCH TECHNOLOGYMANAGEMENT.

RESEARCH/TECHNOLOGY MANAGEMENT

Organizations that create and preserve value in their marketplaces prosper and grow. Those that do not run the risk of disappearing. Creating and preserving value requires continuous technology renewal and leadership capable of responding to new challenges in an environment that is quite different from that of even a half-decade earlier (1).

Because of the breadth of the challenges and the responses required to meet them, the broader terms "technology management" and "research technology management" are used in this article, rather than simply research management or research and development management. By way of comparison, the journal of the Industrial Research Institute, formerly *Research Management*, is now titled *Research•Technology Management*.

Exxon Chemical's first Technology Vice President has explained these distinctions (2):

The role of the Technology Manager is far broader (than that of the R&D manager who develops technologies primarily from internal sources): he or she must see to it that the corporation has available and uses the technologies it needs from whatever source they may be procured... The difference is both subtle and far reaching.

Furthermore, in the 1990s a number of organizations have elevated a senior research and development (R&D) manager to the position of Chief Technology Officer (CTO) in a significant effort to develop a leadership role as the spokesperson for and orchestrator of a technical strategy. This CTO acts as a technical businessperson deeply involved in shaping and implementing corporate strategy (3). By the year 2005, companies may go even further and speak of innovation management.

This article explores some challenges to be faced, the leadership roles demanded, the management issues involved, some best practices used to get results, and some possible futures in research technology management. Although the focus is on industrial technology management in the United States, many of the challenges and issues are present around the world, in government, and in university R&D.

The Challenges Faced

Technology leaders face special challenges, ie, fewer resources, integration with the business, complexity, changing work relationships, and government policies and societal concerns.

Fewer Resources. To assure economic well-being, the investment of $\$175 \times 10^9$ in R&D (1996) in the United States and $\$370 \times 10^9$ worldwide must be managed to maximize return (4). This will require substantial change for over one million scientists, engineers, technicians, and support personnel in 13,000 laboratories across the United States and their counterparts around the world (5).

Investment in U.S. industrial R&D has grown steadily, from $\$5 \times 10^9$ in 1953 to $\$54 \times 10^9$ in 1980, to $\$121 \times 10^9$ in 1995, but the effort in constant dollars peaked, leveled off, and started to decline in the early 1990s (4). This decline is a source of some concern for both the country and individual companies. The concern is heightened as returns from R&D are diminishing. Global competition is reducing margins, new product developments are reducing product life cycles, global information transfer can erode patent protection, and for some fields, the R&D results themselves face diminishing returns. Furthermore, the cost of doing research and development is rising (6). The resulting declining budgets have required some organizational downsizing, a push for greater operating efficiencies, extreme attention to priorities, and a possibly increased focus on lower risk, near-term projects for more rapid payoffs.

Integration With the Business. Within U.S. firms, R&D is becoming more integrated with the business. In 1996 R&D is managed as a system within the larger context of the business (7). Inputs, such as people, facilities, and funds, are processed in the "lab" to produce outputs, such as patents, knowledge, and information for new or improved products, manufacturing processes, services, and businesses. These outputs must be accepted by receivers in the business, such as manufacturing or marketing, in order to be transformed into outcomes, such as cost reductions, new products, new manufacturing processes, and new services (Fig. 1).

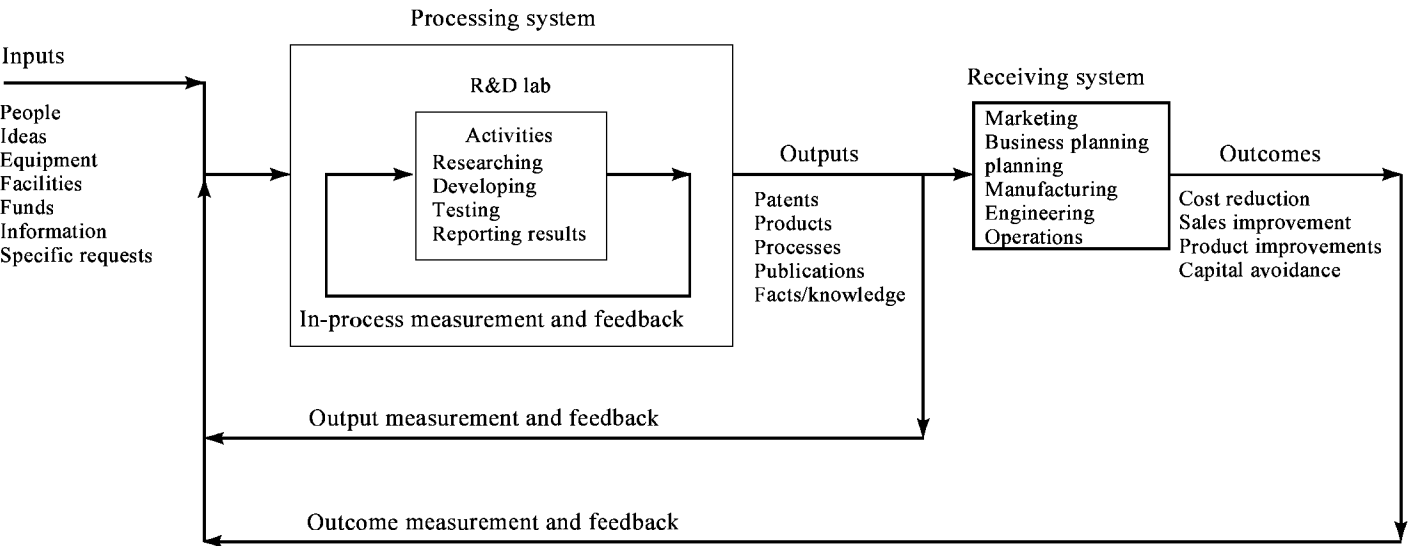


Fig. 1. R&D as a system (7).

Within the larger system, research and especially development processes have become business processes, with teams from marketing, manufacturing, and finance participating from the start (8). In some companies, the R&D function is so highly integrated with the business that formal R&D organizations have vanished; many corporate laboratories have been reduced or disbanded, with the researchers moved into separate business divisions (9,10).

Greater integration of R&D into the business, however, has not necessarily meant a corresponding increase in influence on strategic matters and resource allocation for many U.S. technology leaders. Business strategies often are developed without the essential input of the technical community. In a global benchmarking study involving 244 companies with 80% of the R&D expenditures in the United States, Europe, and Japan, it was found that chief technology officers have board-level membership in almost all Japanese companies, in over half the European companies, but only in 20% of U.S. firms. In addition, U.S. CTOs have far less influence on business-unit technology direction than their Japanese and European counterparts (11). In the extreme case, research facilities may be neglected, technical competencies are at risk, and research can become little more than advanced customer technical service. Some technology leaders report that they have become isolated, feel unfairly judged, and are being held accountable for meeting corporate expectations that they did not help shape (12,13).

Where technology is central to competitive success, such a breakdown in the management of technology can become a serious threat to long-term corporate performance. This situation has created a challenging environment for technology leaders as they struggle to reestablish their credibility and reassert their leadership in the business.

Complexity. Technology management has become more complex with global competition in ever more fragmented markets and technologies. A greater array of inputs from many different disciplines, sources, and locations must be brought together in order to develop new and improved products, processes, services, and applications. The technology organization must also support manufacturing and markets, deal with environmental issues, and fully understand and respond to emerging technologies and scientific advances. All of this has to be done with a speed that surpasses that of the competition.

Almost all of the world's industrial research is conducted by companies from the United States, Japan, and Europe. Companies from these areas have accelerated the global dispersion of their R&D since the mid-1980s (14–16) (Fig. 2). In the twenty-first century, other R&D powers will emerge from India, China, the Asia-Pacific region, and Eastern Europe, unless these areas are disrupted by social, political, and environmental forces (17).

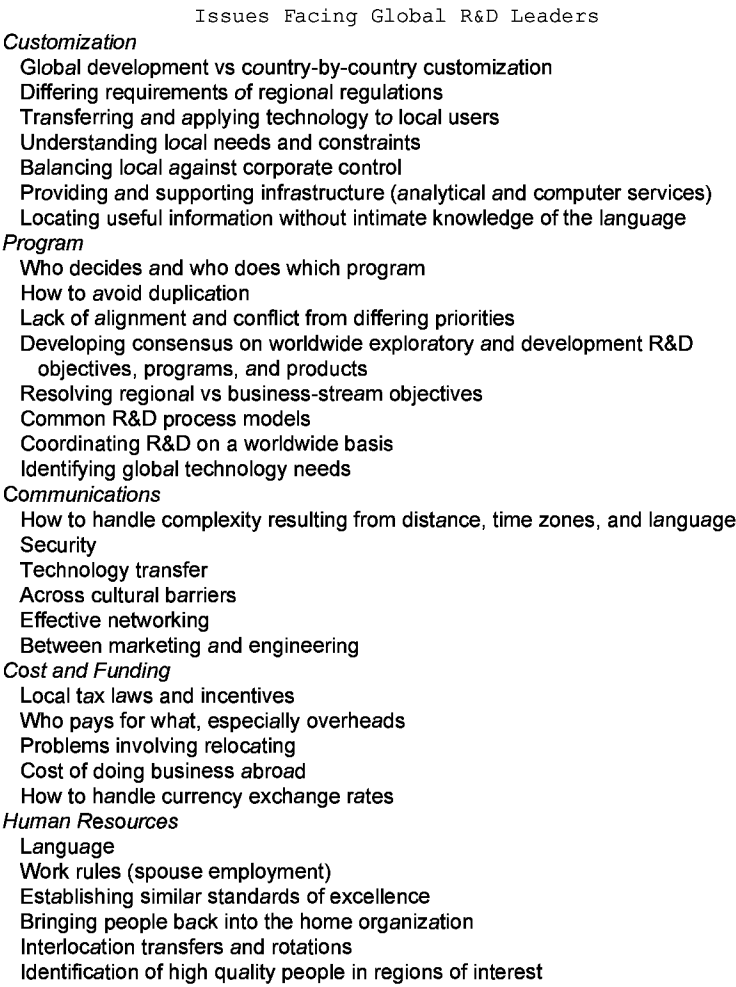


Fig. 2. Issues facing R&D leaders in the face of the global dispersion of R&D (15).

These geographically dispersed efforts require technology leaders who reach beyond traditional borders and create ever more complex technology sourcing arrangements involving licensing (qv) and technology acquisitions, collaborative efforts with suppliers, joint ventures and strategic alliances, consortia and association activities, and relationships with universities, research laboratories, and governmental organizations and laboratories. Technology leaders indicate they have moved from 5–10% dependence on external sourcing in the mid-1980s to 10–20% by the mid-1990s, with the anticipation that the figure could grow to as much as 50% by the year 2001 as organizational and proprietary information barriers are overcome (18). They must further recognize the need to increase the payoff from these sources and to develop an integrated approach when managing both the internal and external R&D. Technology leaders facing tasks with greater scope and dispersion need broader experiences, perspectives, and skills in order to manage effectively in this complex environment.

Changing Work Relationships. The individual doing R&D contributes to the enterprise through a greater variety of working relationships. A few scientists may still make contributions on their own, but it is more common for them to be part of collaborative efforts involving colleagues from around the world or to be on multifunctional teams set up for expediting key product and process developments. The technology leader must not only ensure that effective working relationships are established and maintained in collaborations, teams, and networks, but must also seek new ways to reward and recognize success.

Government Policies and Social Concerns. The creation of economic value through technology development is primarily the responsibility of the private sector. Government has a significant role, however, in creating an environment that fosters such technological innovation. Particularly critical is its funding policy for the national laboratories and academic institutions in the United States that undertake basic research. The government must continue to provide incentives that encourage partnerships between these institutions and industry. The roles of the national laboratories are being scrutinized and their missions redirected to better serve civilian needs; basic research is being questioned and demands are arising to focus more on short-term economic and social issues. The industrial technology community must be an active participant in this debate and in the ensuing decision-making process, and be able to articulate and defend the industrial position on research and technology. This will ensure that government policies encourage a balanced science and technology agenda.

General support for science and technology remains strong in the United States (19,20), but actual financial support for scientific research remains under increased pressure from competing critical needs, including health care, crime prevention, education, pollution control, and national defense. In addition, the decreasing scientific literacy among the U.S. general public, in the national media, and in state and federal government raises concerns over the society's ability to assess and balance technological risks and rewards objectively, and to put them in perspective with other needs.

The technology leader must recognize that in the context in which industrial R&D is conducted, certain societal concerns over the perceived risks can make specific technologies, products, or processes unacceptable in the marketplace and render some technological innovation unfortunately irrelevant.

Leadership Roles

Studies of some extraordinary U.S. successes in high tech markets, eg, GE Medical Systems, GE Plastics, Motorola Communications, and Corning (21,22), have demonstrated the importance of leadership in building competitive advantage through technology. In these successful firms, corporate strategies were shaped around technology-based opportunities and supported by general management. Five conclusions were drawn from the studies. (1) The only way toward market leadership is through a focus on the unique strengths or competencies of the business. (2) This success in the marketplace must come first, followed by a strategic focus. Strategic focus without a unique and sustainable source of advantage is useless. (3) In high tech markets, the only sustainable basis for competitive advantage is technology leadership in a particular area. Strong marketing and distribution reinforce technology leadership, but cannot substitute for it. (4) Technology leadership must be built on both continuous (incremental) and discontinuous (breakthrough) improvements. Neither alone is sufficient. (5) Finally, technology leadership is built by general management decisions driven initially by strategic imperatives and then constrained by financial considerations, rather than by decisions based initially or solely on financial objectives. These conclusions clarify several important roles for technology leaders. They must provide strategic focus, build on strengths, promote innovation, maximize human capital, be a spokesperson for technology, and understand the historical context.

Strategic Focus. As an integral part of the business, the technology leader must see that there is a consistent strategic focus and that technology development is linked to that strategy. For some firms, that strategy is seen in their core ideologies expressed over time, as in the following examples (23).

- 3M Innovation: "Thou shall not kill a new product idea" or "Our real business is solving problems."
- GE Improving the quality of life through technology and innovation; interdependent balance between responsibility to customers, employees, society, and shareholders (no clear hierarchy); individual responsibility and opportunity; honesty and integrity.
- SONY To experience the sheer joy that comes from the advancement, application, and innovation of technology that benefits the general public; to elevate the Japanese culture and national status; to be a pioneer...not following others, but doing the impossible, respecting and encouraging each individual's ability and creativity.

For all firms, technology plans must be closely aligned with and linked to the business plans; they must be linked in a way that provides impact and creates challenges to the business through new options and opportunities. This means that a dual strategic role must be accommodated, one that satisfies both the short-term needs of the individual businesses and the longer-term strategic needs of the company. The technology leader must be willing to challenge the business leaders to disturb the normal equilibrium of business.

Building On Strengths. Technology development must be built on organizational and technological strengths. Technology leaders must understand the basis for these strengths, take responsibility for maintaining them, and ensure that the business recognizes their strategic implications. Leaders must also develop a critical perspective on competition, markets, and customers and place it in a technical context for the business.

The key to creating world-class products constantly, whether companies choose to compete on the basis of speed, service, innovation, value, design, product proliferation, or any combination of these, is to use seven difficult-to-develop capabilities or strengths across all functions (24): (1) create a customer obsession, (2) have a bold strategy or plan, (3) use a multifunctional product-creation process, (4) empower program managers, (5) develop committed teams, (6) marshal the best resources, and (7) speed up the entire process. Competitive advantage can be built upon three different types of strengths all involving technology (25–28): core competencies, best practices, and market access.

Core Competencies. These are sets of skills and capabilities that are difficult to duplicate and have the potential to create entirely new businesses. Technical competencies often refer to sets of key technologies and the kinds of learning embedded in the organization and its people, the competency carriers.

Core technical competencies may be composed of a number of core or key technologies. The competencies in turn can support product families, platforms, or core products, which then support individual products. These products may ultimately be found in a number of forms or shapes. For example, a key technology such as polymer characterization may support a competency in polymer synthesis and architecture, which in turn supports the platform of fluoropolymers and the product family of Teflon (DuPont) fluoropolymer resins that can be found as films, fibers, or in other forms.

One of the important roles of technology leaders requires that they understand how core technical competencies can be used to create opportunities and competitive advantage. To do this, leaders must (1) involve the entire technical community in understanding and assessing the key technologies and areas of science that underlie these competencies; (2) develop and encourage company-wide networks to maintain and support each competency; (3) seek out new business opportunities based on leveraging the core technical competencies or building on individual strengths, and (4) identify logical new competencies that complement or grow out of established ones and that can support growth into new areas.

Eastman Kodak has identified 10 core competencies and developed a process for their management and utilization within the company (29). Similarly, Eaton Corporation selected seven core technical competencies, ranked them in importance, assessed their importance vs the known state-of-the-art for the industry, and developed action plans to extend the life of each (30). Eaton subsequently found the company could bring to market products designed with proven building blocks, thus minimizing risk and the need for additional capital equipment. In addition, the competencies were found to be reservoirs of proprietary advantage that had not previously been put to work.

Best Practices. Technology leaders must provide the direction for improving work processes as well as the work itself. They should seek the best known ways to carry out R&D best practices. There is considerable advice available on best R&D practices.

Market Access. Technology organizations must understand the value chain, ie, the sources of demand for the firm's products and services; the technologies used by customers, learned through a vibrant applications development effort; and the ways in which technological solutions can affect demand. This overall effort requires close contact with the firm's marketing function and with customers in their markets, so that technologies and markets can be matched.

One surprising perspective that emerges is that the selection of customers, not products, becomes the strategic choice for the firm. The starting point for strategy formulation is the definition of distinctive competencies to understand customers and their needs, and to act on those needs (31). This view creates special challenges for the technology leader within a business. How many R&D units feel the responsibility to select the customer base, deciding which customers to serve and which to avoid (32)?

Technology leaders must also understand the parts of the value chain where the firm's competencies are complementary and can create the desired advantage. One tool for this matching and the setting of priorities is the familiarity matrix (33) (Fig. 3). A technical development or entirely new business is placed in a 3 × 3 grid which plots familiarity of the technology embodied in the product or service, ie, known, new but familiar to the firm, or new to the world, against the familiarity of the particular market, ie, known, new but familiar, or new and unfamiliar. This matrix can be used for portfolio analysis or for developing market entry strategies, eg, joint ventures, acquisitions, licensing, or internal ventures. To be effective, therefore, technical people must take responsibility for understanding key market dynamics as well as technologies.

		Product or technology		
		Familiar	New to firm	New to world
Markets	Familiar	Market penetration	Product extension	Product expansion
	New to firm	Market extension	Business extension	Business expansion
	New to world	Market expansion	Business expansion	Entirely new business

Fig. 3. Familiarity matrix, plotting the familiarity of the product's or service's technology against the familiarity of the market involved.

Promoting Innovation. The innovation process involves the search for changes or the causes of changes, and then the analysis and exploitation of the opportunities that are made possible by these changes (34).

Technology leaders must pioneer discontinuous radical changes along with continuous incremental innovations. Radical innovation establishes and periodically renews the competitive advantage; incremental innovations sustain it. Pressures on the leaders are to focus on only the near-term incremental efforts. Technology leaders must, however, build the will within the business leadership to recognize that both are necessary (35).

This dual responsibility can be enabled by the existence of a laboratory that operates with a certain degree of independence within the corporation (3). Such a laboratory has the following critical responsibilities: (1) keep the science base renewed and vibrant, maintaining the base of science experts and

state-of-the-art equipment; (2) leverage competencies, expertise, and resources across the firm in an affordable way, ensuring knowledge flow in shared areas of technical competencies; (3) explore and develop new technologies, competencies, and business options beyond the constraints of individual businesses; (4) maintain a longer-term focus, separate from the day-to-day concerns of the business; and (5) provide access to the world's best hires, technology leaders, consultants, and collaborators.

In an environment in which all R&D is decentralized in the business units, the technology leader's ability to manage this dual role becomes difficult if not impossible.

Technology leaders must maintain a sense of urgency aimed at developing a constant stream of technology results, thus providing the business with options and opportunities to confront the best of competition. This is done by understanding the technology supply chain (Fig. 4), facilitating the flow of information back and forth along this chain, and allocating sufficient resources at each point along it. No piece can be missing if speed of information transfer and speed of delivery are to be increased. How resources are allocated can critically affect productivity. Studies have shown (36) that gains in productivity of individual firms are directly related to the intensity of their investments in R&D, especially when resources are devoted to product and process development; investments in basic research contribute indirectly, but those used in support of technical service can have a negative impact if the resources are taken from product and process development. There are key strategic questions for each link in the chain:

Link in technology supply chain	Key question(s)
science base	Do we have access to the new advances?
market knowledge	How do we maximize this knowledge and test its validity?
invention	Do we have the creative climate, tools, and resources conducive to the creation of inventions? Do we recognize the importance of a proprietary position? Have we developed a patent strategy?
product development	Do we have a quality process for rapid product development and market penetration? Do we have a stream of product improvements? Are we develop-ing common platforms for families of products (37)?
process development	Are we developing improved processes concurrently with our new products? How do we maintain our manufacturing processes at world-class levels (Fig. 5) (38)?
application development	Do we recognize that key inventions are often made here? Do we learn from our lead end users (39)?
customer process development	Do we understand the "technologies of the market-place"? How can we use this knowledge for obtaining a competitive advantage?
plant support	Are we controlling and operating our plants at optimum conditions? Do we really understand our processes as a basis for significant improvement?
customer support	Are the right people supporting our customers? Are they helping to define further opportunities in the marketplace?

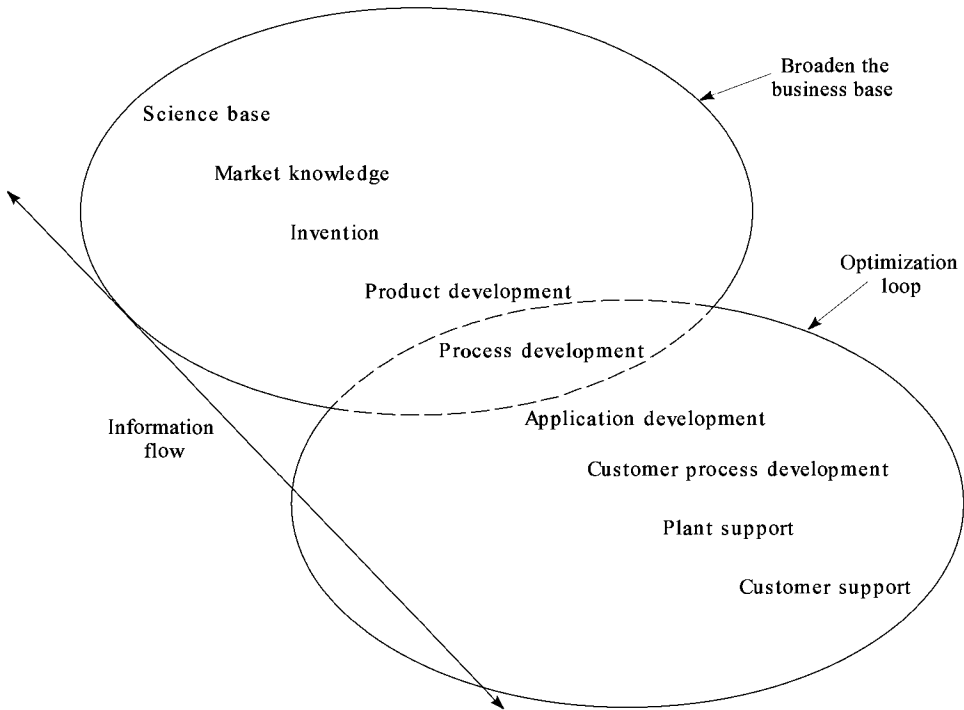


Fig. 4. The technology supply chain, facilitating the flow of information and allocation of resources. This is a necessary tool, but not sufficient in itself in developing a constant stream of technology results.

- Eight-Step Process for Having Best-in-the-Industry Manufacturing Capabilit
1. Identify breakthrough technology options
 2. Develop fundamental and profound process understanding
 3. Capture understanding in process models
 4. Design, build, and operate fully integrated pilot facilities where appropriate
 5. Obtain real-time process data: confirm process models
 6. Design and build the most effective manufacturing facilities
 7. Implement sustainable real-time process control strategies
 8. Fully integrate the manufacturing unit into the business operations

Fig. 5. Eight-step process aimed at maintaining manufacturing processes at world-class levels.

Maximizing Human Capital. Technology leaders must understand, inspire, and guide the scientists and the engineers in recognizing a set of common values, including independence of thought and approach, unwavering commitment to scientific and technical excellence, and a desire for

stability in long-term direction.

These scientists and engineers represent a special challenge to leadership in that the values and motivations may at times be at odds with corporate cultures that emphasize seniority, authority based on hierarchical influence, allegiance to corporate direction, a strict proprietary view of the results of science and technology, and expectations of instantaneous organizational response to changes in direction.

These conflicts require that technology leaders influence technical employees in special ways, eg, by demonstrating their own technical literacy, by encouraging scientists and engineers to participate in guiding the work, by fostering a creative environment, by developing a vision that defines the challenge and relevance of the work at hand, and by encouraging and supporting a certain amount of risk-taking.

What environmental factors stimulate and maintain research productivity? An extensive review of articles and books published over the period 1960–1990 (40) found that a productive research enterprise can be quite fragile, and is dependent on the existence and effective working of 12 individual, organization, and leadership characteristics: (1) clear goals that serve a coordinating function; (2) a research emphasis; (3) a distinctive culture; (4) a positive group climate; (5) assertive, participative governance; (6) decentralized organization; (7) frequent communications; (8) accessible resources, particularly human; (9) sufficient size, age, and diversity of the research group; (10) appropriate rewards; (11) a concentration on recruitment and selection; and (12) leaders with research expertise and skill both in initiating appropriate organizational structures and in using participative management practices. All of these characteristics are interdependent.

Speaking for Technology. The technology leader of the 1990s must be concerned with all aspects of the system from invention through commercialization, as well as with the factors that enable and foster scientific and technological literacy. The leader, therefore, has multiple roles to fill, ie, those of being at once a spokesperson for the value and contributions of technology, a lobbyist for governmental policies that support technology development, and an educator to the public in promoting the need for increased scientific literacy. Three areas are critical: science policy, technology policy, and scientific and technological literacy.

Science Policy. Technology leaders must work to ensure both a long-term national commitment at the federal level to support such basic research (41), and national policies that enable the partnering needed to take advantage of it.

Technology Policy. The federal government must continue to recognize its dual role in support of technology development, both as an enabler through its agencies and national laboratories and as an influential first customer through various departments and agencies. Technology leaders must work to ensure that policies are established that favor innovation and commercialization and that support the complementary roles played with industry in the development of civilian technology.

Scientific and Technological Literacy. There is increasing concern within the business community about the effectiveness of the precollege education system (K–12) in preparing students for an increasingly technically sophisticated workplace. This issue is of special importance to the chemical enterprise, which will find it more difficult to operate effectively in a society that does not understand science and the scientific process (42).

From a business perspective, the education system from K–16 needs to be viewed and managed as a supply chain that establishes the basis for lifelong learning. Increasing the effectiveness of this system will require establishing strong partnerships between all suppliers (precollege through college) and the business community.

Understanding the Historical Context. Technology leaders must understand their corporate and business history and use the understanding as the foundation on which to build the mission and strategic plans. For example, DuPont has a well-documented history of innovation, dating from its beginnings as a manufacturer of black powder (43). Managers familiar with its history recognize the value that technology has brought to the company, the challenges faced with the development and commercialization of new technology, and the periodic waves of innovation that have characterized company growth (35).

The commitment to fundamental research which laid the groundwork for neoprene, nylon, and Teflon (DuPont) dates back to 1926 with the writing of a budget letter (44):

We are including in the central Chemical Department's budget for 1927 an item of \$20,000 to cover what may be called, for want of a better name, pure science or fundamental research work... The sort of work we refer to is work undertaken with the objective of establishing or discovering new scientific facts... The volume of fundamental work is rapidly losing ground as compared with the volume of applied research. In other words, applied research is facing a shortage of its principal raw materials.

In 1950, the then-president of DuPont gave the prescription for managing R&D (45):

Provide a broad commercial base...leverage research across diverse business lines. Pick your problems with great care and judgment. Be patient; don't evaluate results after one year, but wait as long as five or six. Know when to quit a problem; done largely by intuition, but we prefer to call it intelligence.

Success, whether achieved in the 1950s or 1990s, involves extremely credible technical leaders who have gained the confidence of the business community by delivering technology with significant business impact.

Managing the Technology Function

Industrial technology leadership involves setting and influencing the direction and course for technology in the firm; management involves pursuing that direction in developing and then implementing specific courses of action.

More specifically, technology management involves five functions: (1) setting the context, ie, positioning R&D as part of the business system; (2) planning; (3) organizing; (4) motivating; and then (5) assessing and improving. Each of these functions must be directed toward effective performance at each of three levels, ie, the entire organization, various work processes, and individual efforts (Fig. 6) (46).

		Performance Needs :				
		Context	Goals	Structure	Motivation	Improvement
Levels of Performance	R&D organization	Is R&D's role in the business understood?	Are technology plans developed and broadly understood?	Are R&D resources organized in the best way to carry out its mission?	Is the organizational climate assessed?	Does the organization have measures for success? Does the organization learn? Audits in place?
	R&D work processes	Are R&D practices linked to business practices?	Are work processes those needed to implement the plans?	Are work processes defined, disciplined, and structured?	Is the human side of work practices understood?	Are there metrics related to the goals/results of the work processes?
	Scientist/ engineer effort	Are global technology networks and collaborations fostered?	Are individual objectives supporting the plan?	Is individual work organized? Are teams adequately chartered?	Are individual motivation factors understood and acted upon?	Are researchers becoming active learners?

Fig. 6. Managing technology: some key questions. Performance needs are examined for each of the five functions of technology management, at three different levels within the business.

Setting the Context. A desired evolution of the management of industrial R&D to third-generation R&D (47) is characterized by the building of partnerships within the firm, ie, bringing R&D and the other business functions together; creating effective R&D processes linked to other business processes; setting R&D technology strategy as part of corporate and business strategies; and developing a shared responsibility for success. This is not an easy task.

A key responsibility for technology management is positioning the R&D technology function in the context of business, ie, clarifying the mission and role of technology and establishing the necessary understanding in the entire business, especially in helping the business leaders to understand the need for the successful management of risk by R&D. Opportunities for growth are created by understanding the risks involved within the context of the firm's strengths and knowledge, and then managing the risks in a disciplined way.

Planning. Through planning, technology managers establish objectives, develop the rationale for these objectives, and estimate the resources required over time to succeed at the strategic business, R&D organizational, and project levels.

Two primary barriers to effective planning are the general reluctance of technology managers and their organizations to plan and the lack of clear business plans or strategies, especially ones that include sufficiently well-defined time horizons. The process of planning should be distinguished from the plan itself. Planning is an analytical and strategy-setting process. The plan itself states the conclusions derived from the process. Of the two, the process is the more important (48). This process should be interactive, iterative, and incremental (49). *Interactive* means it should be done by teams of decision-makers from both the business and the technology organization. *Iterative* means it is not a linear, simple process. Higher level decisions are often provisional and are revised after more detailed analysis. *Incremental* means building on current plans.

Technology planning is valuable when it provides R&D with a clear direction linked to the strategic needs of the business, helps balance the short- and the long-term focuses, and helps the organization stay on-course. Planning forces R&D managers to set priorities, to give clear guidance to scientists and engineers as to why R&D is needed, and to evaluate R&D proposals. The process facilitates human and capital resource planning and allocation, forces the organization to evaluate the future potential of technologies in a changing world, provides a foundation upon which to make choices when selecting technologies to pursue, and helps organizations to explore ways of exploiting these competencies in order to build platforms for new business growth.

Important characteristics of effective planning processes include the participation of a wide range of technically knowledgeable individuals, timing that is "evergreen," adequate communication of the plans both inside and outside the organization, and periodic monitoring of progress vs the plan, by the business as well as by the R&D technology organization.

Technology managers should include the following elements in the plan itself: the critical business issues; the nature of markets and customers; the thrusts of traditional and potential competitors; the human resource issues, ie, the skills and abilities needed to carry out the plan; alternatives and contingencies to support the plan; a patent and intellectual property strategy; and the supporting facilities and resource plan. The most frequently overlooked elements are human resource issues and patent strategies (50).

A number of special approaches can be used. For example, Motorola developed two technology roadmap processes (51), one for dealing with a single technology and one for dealing with a product line. The single-technology roadmap provides an objective evaluation of Motorola's capabilities in the technology, a comparison of those capabilities with the opposing capabilities of competitors in the 1990s and in the future, and a forecast of progress in that technology. The product-line roadmap includes, in addition, technological requirements for future products and possible technology sources; the impact of technology on product properties and quality; and a minority point of view, ie, other perspectives, alternatives, and or technologies that might have been ignored.

Every business day there are more than 5000 research papers published and 1000 new patent documents issued. Analysis of these activities (52) can provide comparative indications of the impact, speed, science linkage, and technology linkage to firms around the world, adding an important element to strategic business and technology planning.

Strategy tables and influence diagrams can also be important management methodologies for coherent as well as shared planning (53). Strategy tables list alternative product concepts, design options, as well as technical and market alternatives.

Influence diagrams can be used with the strategy tables to identify factors or variables that create uncertainty in assessing the value of various options. Issues can be qualitatively prioritized in selecting the best path to generate value for the business.

In dealing with future uncertainties, Royal Dutch Shell pioneered Scenario planning (54,55). Alternative assumptions for future developments can be combined under this approach in various ways to give a number of consistent possible outcomes (56) and provide a basis for both actions and reactions. The approach has rewarded Shell handsomely.

R&D and technology plans can also be analyzed with widely used portfolio analysis (47), which provides visual correlations between variables such as potential value, probability of success, time-to-completion, technology maturity, and project cost as a means of optimizing the planned effort. The analyses can be used to redirect effort and reallocate resources as a project progresses.

Finally, a constraint analysis has been developed to screen new opportunities (57). This analysis separately evaluates intrinsic business attractiveness factors and the fit factors required for commercial success.

Business attractiveness	Business fit factors
sales	capital availability
profit potential	manufacturing-production
growth potential	marketing-distribution
competition	technical competence
distribution risk	access to components (raw materials)
restructuring potential	management as champions
political	
social factors	

Each factor is evaluated on the basis of a score of 0–10. A total score of 80 points or higher statistically has resulted in eight or nine successes out of ten instances. Below 80 points, the success-to-failure ratio for new development drops off sharply. Such analyses can be made not only initially but periodically during an R&D project and as the business develops.

Organizing. Organization of the R&D effort involves seeking out, bringing together, integrating, and structuring the company and its resources, the work processes, and the effort of individuals to accomplish the set goals and objectives.

Today's technology manager faces the special challenge of orchestrating a vast array of resources, including many that can be located outside the manager's organizational unit. For example, concern over depletion of the earth's protective ozone layer led to the decision to phase out the manufacture and sale of chlorofluorocarbons. In a matter of five years, the Freon refrigerants business in DuPont was remade into an entirely new Suva refrigerants business. Technology development valued at $\$400 \times 10^6$ in capital and R&D resulted in the creation of seven new manufacturing facilities and dozens of products grouped in five families to replace the six in the Freon (DuPont) business, and also involved the filing of 400 patents. This would not have been accomplished had not the R&D manager created a network of supporting resources that included universities, institutes, customers, and industry consortia. Basic data for synthesis and manufacturing as well as toxicity and property information came from over two dozen technical resources around the world. The result was the successful re-creation of a $\$500 \times 10^6$ to $\$1 \times 10^9$ business (58).

Technology managers are devoting more attention to developing and improving their organizational work processes and the particular practices used in these processes, which range from recruiting, to budgeting, to all the interconnected innovation processes. Furthermore, technology managers are harnessing technologies and entirely new tools to increase the productivity of R&D work. Computational chemistry and modeling eliminate some laboratory work altogether, new analytical tools arrive at answers faster, experimental equipment ranging from miniature computerized semiworks to fully automated laboratories feeds information directly to databases, and readily available on-line information sources prepare the researcher more effectively.

Networking within the technical community represents an additional general process which needs to be fostered by technology managers. Technical networks are powerful forces for nourishing and applying key technologies. They are important technology transfer vehicles. Networks in 3M, Hughes Electronics, DuPont, and Martin-Marietta have been described as ways for linking technology and technologists, thus promoting information flow along the technology supply chain, increasing efficiency, and improving technical competitive advantage (59,60). Technology councils have proven to be effective networks in a number of companies, linking technology leaders and addressing issues related to contemporary technology management. Scientific councils, fellows forums, or company academies provide ways for top scientists and engineers to study issues, monitor emerging technologies, and increase the business impact of the technical community in the firm.

A study conducted at Bell Labs revealed that the real difference between star and average workers was not IQ but the way top performers do their job. One of nine key work strategies was networking: getting direct and immediate access to co-workers with technical expertise and sharing one's own knowledge with those who need it (61).

Motivating. Motivation within technical organizations involves not only energizing individuals but lifting the morale and spirit of the organization as a whole.

Technology leaders must understand that many values and expectations of the technical professional are unique and differ from those typical of others in the workplace (62,63). Moreover, technical organizations have unique cultures developed in the scientific tradition and characterized by collegiality, rigorous exploration, forthright challenges to one another, continuous development, and a high degree of skepticism. These values and expectations may be described as follows.

Desire to Be Independent Yet Collaborate. Technical professionals are achievement-oriented and derive motivation from the work itself. This often means a need for increased participation in decision-making, as well as a natural skepticism and resistance to rapid changes in direction and tight control by managers. Collaboration and the desire for some form of teamwork also are a part of the professional's nature.

Need for Achievement. Professionals are often motivated by a need to accomplish complex tasks requiring high levels of skill and effort. Strong commitment is generated when the work is seen as exciting, meaningful, and intellectually challenging.

Fear of Obsolescence. Demotivation and declining performance occurs when the technical professionals perceive that their skills are being eroded or not utilized.

Professional Identification. Many technical professionals identify as strongly with their professions as with their company. Conflicts arise when organizational goals are not aligned with professional goals and the promotion of technical excellence.

Participation in Developing the Mission and Establishing Goals. Technical professionals are often more resistant to mandated goals than are others in the organization. Participation in goal-setting is integral to their job satisfaction; alignment of individual and organizational goals is important for sustaining motivation. Once committed to the goals, technical professionals often set very high performance standards involving an intense personal commitment; sudden changes in objectives or goals can be particularly upsetting and demotivating. They are, furthermore, very interested in the needs of the business and in how they can contribute to business success. It may even be said that they thirst for this type of knowledge.

Collegial Support, Stimulation, and Sharing. Technical professionals value collegial support and an environment that makes use of a diversity of knowledge and experience. Productivity of scientists has been correlated with the number of collegial collaborations (64).

Success Factors. These values and expectations have been reflected in periodic morale satisfaction surveys taken within several technical organizations (65). Morale was found to be highest when (1) R&D leaders were visible and available, communicated the direction and mission, explained the reasons for tough decisions, reviewed professional advancement, understood the work being done and the impact of decisions, led the effort to balance resources with the work at hand, and provided a relatively secure environment; (2) R&D managers involved individuals in decisions, recognized accomplishments, developed meaningful reward and recognition practices, studied priorities, linked work to meaningful needs, secured the resources, helped the professionals to focus on the exciting work and get it done, and supported continuing skill development; and (3) professionals gave credit and support to co-workers, accepted change, looked to make improvements within each one's control, focused on the work, and sought meaningful accomplishments.

The technical workforce has become more diverse. Technology managers need to educate themselves about the changing nature of the workforce, the implications for the future, and the opportunities made possible by a more heterogeneous organization (66).

The degree to which technology managers use all of this knowledge to guide their decisions can determine the extent to which key objectives are met.

Assessing and Improving. As the costs of R&D rise and as R&D becomes more integrated into the business, a number of constituencies or stakeholders are questioning the value, the effectiveness, and the relative productivity of R&D. Technology leaders find that they must define measures for R&D not only to measure effectiveness, but also to justify R&D efforts to others. These are not simple tasks. Table 1 summarizes a way to categorize some of the metrics that have been proposed and the constituency that will find the greatest utility for each category (67).

Table 1. Measures for R&D^a

Category	Primary constituent	Examples of metrics
value creation	financial community, CEO	<i>Output</i>
		financial return R&D cost
		sales of new products and processes
		projected value of R&D pipeline
		break-even time
		manufacturing cost

portfolio management; integration with the business	business management R&D leaders	product quality profit margin market share
		<i>Strategy</i> strategic alignment investment in key and pacing technologies distribution by objective
asset value of technology R&D	R&D leadersR&D staff	distribution by project type distribution by impact time distribution by core competency distribution by internal or external activity number of ways technology is exploited number of product definitions having business approval number of business commitments at various R&D pipeline events percentage of funding by businesses technology transfer to manufacturing use of cross-functional teams
		<i>Foundations</i> customer rating of product technology benefits, response time to competitor moves, current investment in technology, quality of personnel, development cycle time, customer rating of technical capability, number and quality of patents, sales under patents, peer evaluation, customer satisfaction, rate and number of development pipeline milestones achieved, customer contact time, specific develop-ment pipeline activities per R&D dollars spent, reports, publications, etc, efficiency of internal process, employee morale, core technical competency, delayed stage skills, cost vs budget, quality of process of data assessment, probability of success, technology planning, environmental management in R&D, use of information technology in R&D, idea generation and creativity, people development, intellectual property management, idea generation, and R&D climate

^a Ref. 67.

Technology managers have also been given frameworks for assessing their technical function's mission, objectives, strategic plan, key policies, and R&D activities. One framework provides indicators for managers to position their organizations and work practices in four stages of development (68). For example, for quality assurance (qv) the stages can be described as follows.

Stage 1	Stage 2
no quality measures finger-pointing	controlled output some design quality measures problems fixed as they arise
Stage 3	Stage 4
checkpoints for in-process quality control during development process	detailed process models and metrics for all routine technical operations; continuous improvement
root cause analysis of quality problems focus on satisfying internal and external customer needs	quality function deployment (QFD)-type methodologies used routinely to assure customer perspective

Another framework (69,70) highlights 10 R&D activities and six levels of performance:

R&D activities	Levels of performance
selecting R&D projects planning and managing projects generating new product ideas	issue is not recognized
maintaining the quality of the R&D process and methods motivating technical people establishing cross-disciplinary terms coordinating R&D and marketing transferring technology to marketing fostering collaboration between R&D and finance linking R&D to business planning	initial efforts at addressing the issue are underway
	right skills are in place
	appropriate methods are used
	responsibilities are clarified
	continuous improvement is underway

In using such assessments, the technology manager needs to answer two important questions: (1) How do I improve our work processes? (2) What processes or practices should have the highest priority for improvement?

Answers to the first question can be illustrated by giving some examples (see Table 1). Eastman Chemical reengineered its innovation process and doubled the value of its R&D portfolio (71–74). A team at Eastman was asked to provide (1) an assessment of the then-current innovation process, (2) a vision of the ideal process, (3) a flow chart of the modified process, (4) measures of the process, and (5) key roles and responsibilities. The team identified four main subprocesses: needs identification, concept development, implementation, and market development.

The detailed analysis, involving many respondents inside and outside the company, led to changes in the overall innovation process, ie, a company-wide priority system for innovation projects, measures of innovation for each functional and business area, training and supportive management systems for project managers, informal multifunctional teams in concept development and market development, a structured needs identification process, and appointment of a process steward to monitor the innovation process, measure how it functions, and coordinate innovation projects.

In other efforts, over 100 firms in the United States and Europe have adopted a specific structured work process for product process development, called product process and cycle time excellence (PACE). In effect, PACE makes new product and new chemical process development a business-driven,

not R&D-driven, process, one involving phase reviews where key decisions on funding, proceeding, or terminating are made by a product approval team from the business (75). Other important features of the process include a core multifunctional team that manages the specific product or process development project, a structured development methodology involving well-understood stage gates and milestones with a detailed schedule, and a work process engineer who sees that this process runs well and seeks ways to improve the process and the related decision-making. This process, which forces complete resourcing and prompt decisions based on well-understood criteria, has reduced overall development time by 30–50% for most users.

To answer the second question, on setting priorities for improvement, the technology manager has many sources of advice. For example, the statement, "The single greatest challenge facing managers...is to raise the productivity of knowledge and service workers...(this) will determine the quality of life in every industrialized nation... In knowledge and service work, working smarter is the only key" (76). But in R&D, what is really meant by "working smarter"? Once again advice is not in short supply. Technology managers can find courses, books (77–84), articles, and long lists of tips (85–91).

In the decade from 1985 to 1995, numerous broad-based studies showed general agreement on a limited set of best R&D practices that have the greatest positive impact (92–98). Taken together, these studies provide a priority listing of the top nine R&D processes and practices.

- (1) *Setting integrated technology strategies, plans, and objectives.* Technology strategies and plans are aligned with business strategies and plans and are well understood across all functions. A common language within the business is used for communicating R&D efforts, objectives, and potential results.
- (2) *Identifying customer needs.* A disciplined process allows for the spotting of needs, understanding what the customer will value, knowing the value that each market segment places on each product property, and understanding how competitors might address the needs.
- (3) *Identifying technical possibilities.* The organization is capable of identifying alternatives in meeting customer needs; solving problems; defining technical projects to meet these needs; characterizing technology, ie, how and why a product works as it does; and identifying technology boundaries, ie, the mechanisms that will limit ultimate technical performance.
- (4) *Selecting programs.* The business has a goal-setting process that collectively develops clear, mutually agreed upon, strategically evaluated project objectives with well-understood expectations that are then clearly communicated. Project selection criteria are clearly linked to business strategy. A prioritization process maintains a backlog of ideas and opportunities.
- (5) *Assuring technical excellence.* Processes are in place to secure and maintain technical excellence (99) among the firm's scientists, engineers, and managers (Fig. 7). The long-term R&D strategy includes the needed personnel development processes.

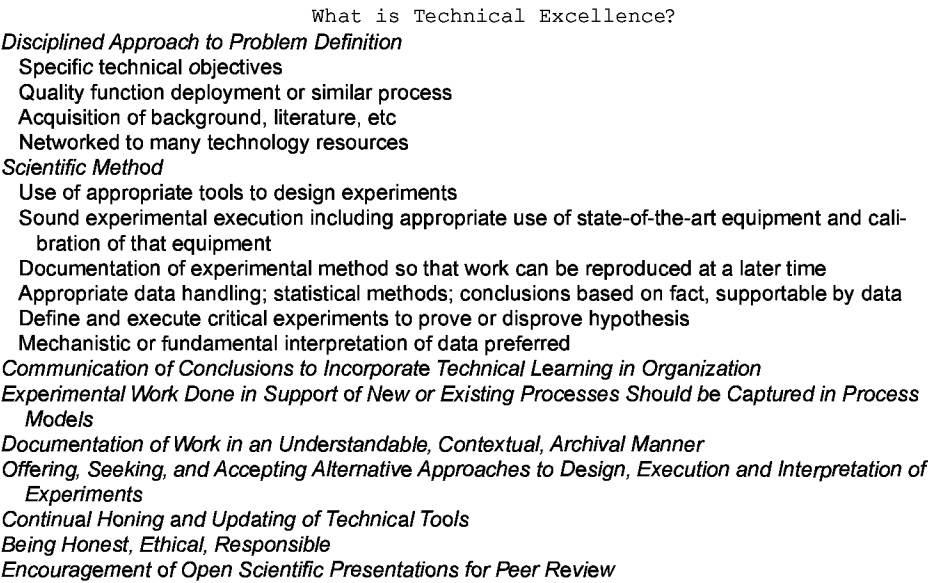


Fig. 7. Elements needed to secure and maintain technical excellence among the firm's scientists, engineers, and managers.

- (6) *Planning, staffing, and terminating projects.* Technology managers have the experience to predict what can be accomplished under certain time and cost constraints; have the authority to pull together the needed set of technical and managerial resources to get the job done; have the knowledge upon which to reach a decision that a project is completed, no longer relevant, or unlikely to succeed, and then have the ability to implement that decision. The project plans assure that the most significant uncertainties or assumptions are addressed as early as possible.
- (7) *Measuring effectiveness.* Results are measured against technology and business objectives; the metrics chosen determine the proper behavior and actions.
- (8) *Understanding core technologies.* A process is in place for defining, maintaining, and creating core technologies and then integrating them into longer-term technology and business plans.
- (9) *Maintaining external awareness.* Beyond understanding customer needs, a set process monitors external threats and opportunities from a variety of sources.

In improving these work practices, technology managers are faced with the somewhat new challenge of learning to manage processes rather than organizations and learning how much structure is put onto the process. If too much structure is imposed, then bureaucracy reigns: there are too many policies and measures, and the discipline is avoided or circumvented in order to get the work done. Given too little structure, however, a free-for-all is present: steps are not repeatable, and researchers are confused and do not understand the process or work at hand (24). Microsoft for example has developed a semistructured product development process (100) that focuses creativity by evolving product features as resources are constrained.

The innovation and product development processes continue to be studied, success stories are highlighted, and technology managers try to improve their work processes. Yet, little improvement has been made since the 1970s in raising the overall new product success rate of companies as a whole. For most substantially new industrial innovations, including industrial chemicals, it takes 3000 unwritten ideas to come up with 300 ideas leading to simple experiments and or the filing of a patent. About nine of these projects make it to significant, larger development efforts, and four of them advance to major developments. Of these, ~1.7 are commercially launched, resulting in one commercial success (101).

Most technology managers realize the challenge in these figures. The development of best practices in their organizations, while necessary, is not sufficient. There is no perfect system for managing R&D precisely. Key intangibles within the organization require judgment and intuition. These skills are gained only by experience and can greatly differentiate one firm from another.

Technology Management in the Twenty-First Century

Considering "What lies ahead?" may be done in three sections (102).

Current Trends. Many current trends will most likely intensify. Company resources devoted to R&D will grow increasingly precious. Those involved in the R&D process will have to find even more ways to work smarter, lower costs, and get more done with less (103). "How much R&D?" will be asked on a regular basis.

The value of R&D vs other business needs will continue to be questioned, and the pressure will increase for near-term results where the value is

readily apparent. Those involved in R&D will need to be even more articulate and will have to find additional ways to measure and demonstrate that value and continued effort will lead to improving success rates.

The R&D function itself will be organized from company to company in very different ways. In some firms, the responsibility for innovation will be broadly shared and a distinct R&D unit may cease to exist. In others, technical developments from an R&D lab will be the principal source of innovation for the business. In still others, the primary task of the R&D or technology group will be to obtain technology from outside the firm. Those involved in R&D will need to make certain that their knowledge and talents impact the business irrespective of their precise function. The innovation process will become less linear (Fig. 8) as more feedback loops, technology transfer, and cooperative efforts are involved.

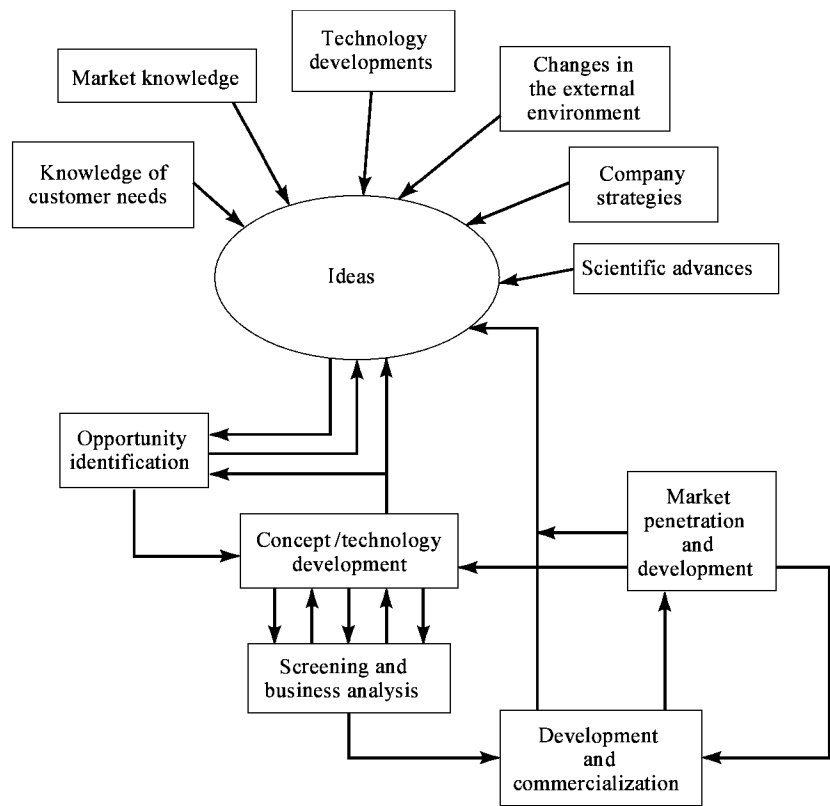


Fig. 8. The innovation process model seen as a series of feedback loops involving technology transfer and the merging of cooperative efforts on a global scale.

Research and development resources will become more dispersed, not only within the firm but also around the globe (104). Managing technology will be a more complex task. Those doing the managing will need to draw on entirely new resources and newer information technologies to make possible additional ways for individuals and groups to work and collaborate with one another while being miles or continents apart.

Various constituencies in society will continue to assert that some technology-based innovations present undue risks. Technology managers will be spending a greater amount of time and effort demonstrating where risks are negligible and manageable, and further where they are far outweighed by the benefits.

New Developments. Current trends can give rise to several developments. Less R&D may be done in what are presently the research-intensive industries, but more R&D will be done in the nonmanufacturing or service industries, building on the newer technologies and using sciences and disciplines other than chemistry, physics, and biology, such as economics, psychology, history, archaeology, and anthropology. Present R&D (ca 1996) in financial services, banking, investing, and transportation will spread to other service industries. The output of this R&D, in turn, will provide further opportunities for the present research-intensive industries.

"Virtual" R&D organizations will be at the heart of collaborative business arrangements formed to meet pressing and complex needs by harnessing a wide variety of technical skills and competencies (105). Such arrangements will be very fluid, dissolving once the task is accomplished and making way for still further alliances. More and more needs will be met by the fusion of several technologies (106). Because of the shortening of cycle times, all members of a value chain will have to be more open and seek collaborations to create value.

Increased ease of communication and information technology and newer ways to conduct R&D may again permit the individual scientist or engineer to function as the program manager and be in effect an entire research or development team. Increased attention will be given to the capture and retention of knowledge by both the individual and the organization as a whole.

Ways will indeed be found to use newer technologies to lower the cost of producing R&D results, including the use of highly sophisticated modeling and simulation to avoid some laboratory, pilot-plant, and applications research altogether.

Unchangeables. Some things, however, will not change. The basic reason for being, for industrial R&D, will remain as it has been since the early 1900s. At the core of this mission was, is, and will continue to be the necessity to understand customer needs and develop solutions linked to business strategies. The fostering of technical excellence will separate the "winners" from "losers." Creativity will be at the heart of research and development processes (107). The continued development of critical science-based technologies will generate new ways for meeting existing needs and create new opportunities for emerging needs, and technology leaders who are articulate, courageous, and who can deal successfully with change will continue to be needed as an integral part of U.S. businesses.

BIBLIOGRAPHY

"Research Management" in *ECT* 2nd ed., Suppl. Vol., pp. 855–875, C. A. Stokes, Consultant; in *ECT* 3rd ed., Vol. 20, pp. 179–196, by A. M. Beuche and L. W. Steele, General Electric Co.

1. H. I. Fushfeld, *Industry's Future, Changing Patterns of Industrial Research*, American Chemical Society, Washington, D.C., 1994, pp. 305–346.
2. J. F. Mathis, *Res. Technol. Mgmt.*, 35 (Jan.–Feb. 1992).
3. W. W. Lewis and L. L. Linden, *Sloan Mgmt. Rev.*, 57–67 (Summer 1990).
4. *Industrial Research and Development Facts*, Industrial Research Institute, Washington, D.C., 1995.
5. *National Patterns of R&D Resources: 1994 (NSF 95-304)*, National Science Foundation,
6. *The Future of America's Research Intensive Industries*, Institute for the Future, Menlo Park, Calif., Aug. 1995, p. 8.
7. M. G. Brown and R. A. Svenson, *Res. Technol. Mgmt.*, 67–71 (July–Aug. 1988).

8. P. F. Drucker, *Wall St. J.*, 20 (Feb. 10, 1988).
9. E. Corcoran, *Res. Technol. Mgmt.*, 14–20 (July–Aug. 1994).
10. E. Corcoran, *Sci. Am.*, 103–110 (June 1992).
11. E. B. Roberts, *Res. Technol. Mgmt.*, 44–56 (Jan.–Feb. 1996).
12. B. Uttal, A. Kantrow, L. H. Linden, S. B. Stock, *Res. Technol. Mgmt.*, 15–24 (May–June 1992).
13. P. R. Bridenbaugh, *Res. Technol. Mgmt.*, 27–33 (Nov.–Dec. 1992).
14. L. C. Krogh, *Res. Technol. Mgmt.*, 25–28 (July–Aug. 1994).
15. E. R. Gilmont, *Res. Technol. Mgmt.*, 7 (May–June 1995).
16. D. H. Dalton and M. G. Serapio, Jr., *Globalizing Industrial Research and Development*, Office of Technology Policy, Washington, D.C., 1995.
17. R. Kaplan, *Atlantic Monthly*, 44–76 (Feb. 1994).
18. R. S. Jonash, *Prism*, Arthur D. Little, Cambridge, Mass., First Quarter 1993, pp. 23–35.
19. *Science and Engineering Indicators*, National Science Foundation, Washington, D.C., 1993.
20. Ref. 6, pp. 23–24.
21. J. G. Morone, *Winning in High Tech Markets*, Harvard Business School Press, Boston, Mass., 1993.
22. J. G. Morone, *Res. Technol. Mgmt.*, 16–25 (Mar.–Apr. 1993).
23. J. C. Collins and J. I. Porras, *Built To Last: Successful Habits of Visionary Companies*, Harper Business, New York, 1994.
24. J. P. Deschamps and P. R. Nayak, *Product Juggernauts: How Companies Mobilize to Generate a Stream of Market Winners*, Harvard Business School Press, Boston, Mass., 1995.
25. G. Hamel and C. K. Prahalad, *Harvard Bus. Rev.*, 63–76 (May–June 1989).
26. C. K. Prahalad and G. Hamel, *Harvard Bus. Rev.*, 79–91 (May–June 1990).
27. G. Hamel and C. K. Prahalad, *Harvard Bus. Rev.*, 81–92 (July–Aug. 1991).
28. G. Hamel and C. K. Prahalad, *Competing For The Future*, Harvard Business School Press, Boston, Mass., 1994.
29. E. P. Przbylowicz and T. W. Faulkner, *Res. Technol. Mgmt.*, 31–38 (Jan.–Feb. 1993).
30. W. J. Walsh, in R. Szakonyi, ed., *Technology Management-3*, Auerbach Publications, Boston, Mass., 1994, pp. 3–15.
31. F. E. Webster, Jr., *Market-Driven Management: Using the New Marketing Concept to Create a Customer-Oriented Company*, John Wiley & Sons, Inc., New York, 1994.
32. P. M. Norling, *Res. Tech. Mgmt.*, 60 (Jan.–Feb. 1996).
33. E. B. Roberts and C. A. Berry, *Sloan Mgmt. Rev.* **26**(3) (Spring 1985).
34. P. F. Drucker, *Innovation and Entrepreneurship: Practice and Principles*, Harper & Row, New York, 1985.
35. J. A. Miller, *Res. Technol. Mgmt.*, 1996, in press.
36. A. S. Bean, *Res. Technol. Mgmt.*, 25 (Jan.–Feb. 1995).
37. J. M. Utterbach and M. H. Meier, "Reducing Cycle Times", *Industrial Research Institute Fall Meeting*, Williamsburg, Va., Oct. 23–26, 1994.
38. J. A. Trainham, "Modeling: A Strategic Tool for Gaining Competitive Advantage", *AspenWorld Conference*, Boston, Mass., Nov. 7, 1994.
39. E. von Hippel, *The Sources of Innovation*, Oxford University Press, New York, 1988.
40. C. L. Bland and M. T. Ruffion VI, *Academ. Med.* **67**(6), 385–397 (June 1992).
41. J. A. Armstrong, "Is Basic Research a Luxury?" *Karl Taylor Compton Lectures*, Massachusetts Institute of Technology, Cambridge, Mass., Oct. 13, 1993.
42. H. E. Simmons, 1994 Priestley Medal Address, *C&E News*, 27–31 (Mar. 14, 1994).
43. D. A. Hounshell and J. K. Smith, Jr., *Science and Corporate Strategy: DuPont R&D, 1902–1980*, Cambridge University Press, Cambridge, U.K., 1988.
44. Ref. 43, p. 223.
45. Ref. 43, pp. 360–361.
46. G. A. Rummler and A. P. Brache, *Improving Performance: How to Manage the White Space on the Organization Chart*, Jossey-Bass, San Francisco, Calif., 1990, p. 19.
47. P. A. Roussel, K. N. Saad, and J. J. Erickson, *Third Generation R&D*, Harvard Business School Press, Boston, Mass., 1991.
48. J. R. Champion, in R. Szakonyi, ed., *Technology Management-2*, Auerbach Publications, Boston, Mass., 1993.
49. A. R. Fusfeld, *Strategic Management of Technology*, DuPont Co., Wilmington, Del., 1993.
50. J. Reagan and J. Tancredi, *Integrating Technology and Business Planning in IRI Companies* (video and Facilitator's notes), Industrial Research Institute, Washington, D.C., 1995.
51. C. H. Willyard and C. W. McClees, *Res. Technol. Mgmt.*, 13–19 (Sept.–Oct. 1987).
52. F. Narin, M. Albert, and V. Smith, *Sci. Pub. Pol.*, 369–381 (Dec. 1992).
53. M. M. Menke, *Res. Technol. Mgmt.*, 25–32 (Sept.–Oct. 1994).
54. P. Wack, *Harvard Bus. Rev.*, 73–89 (Sept.–Oct. 1985).
55. P. Wack, *Harvard Bus. Rev.*, 139–150 (Nov.–Dec. 1985).
56. H. Geschka, M. Verhagen, and B. Winckler-Rub, in R. Szakonyi, ed., *Technology Management-2*, Auerbach Publications, Boston, Mass., 1993, pp. 399–416.
57. D. B. Merrifield, *Res. Technol. Mgmt.*, 14–18 (July–Aug. 1993).
58. P. M. Norling, "Accessing Technology from External Sources: Using External Sources for CFC Alternatives," *Industrial Research Institute Roundtable*, Washington, D.C., July 28, 1994.
59. A. N. Chester, *Res. Technol. Mgmt.*, 25–32 (Jan.–Feb. 1994).
60. P. M. Norling, *Res. Technol. Mgmt.*, 42–48 (Jan.–Feb. 1996).
61. R. Kelley and Janet Caplan, *Harvard Bus. Rev.*, 128–139 (July–Aug. 1993).
62. M. K. Badawy, *Res. Technol. Mgmt.*, 19–35 (Sept.–Oct. 1988).
63. *How to Lead Today's Technical Professional: Foundation Document for Technical Leadership Training*, Mohr Development Inc., Stanford, Conn., 1989.
64. D. Pelz and F. M. Andrews, *Scientists in Organizations*, rev. ed., University of Michigan Press, Ann Arbor, 1976.
65. P. M. Norling, technical data, DuPont Co., Wilmington, Del., 1991–1996.
66. G. G. Gordon, N. DiTomaso, and G. F. Farris, *Res. Technol. Mgmt.*, 18–23 (Jan.–Feb. 1991).
67. J. W. Tipping, E. Zeffren, and A. R. Fusfeld, *Res. Technol. Mgmt.*, 22–39 (Sept.–Oct. 1995). A searchable database with more than 50 metrics is available from the Industrial Research Institute, Ste. 1100, 1550 M St., N.W., Washington, D.C. 20005-1708.
68. P. S. Adler, D. W. McDonald, and F. MacDonald, *Sloan Mgmt. Rev.*, 19–37 (Winter 1992).
69. R. Szakonyi, *Res. Technol. Mgmt.*, 27–32 (Mar.–Apr. 1994).
70. R. Szakonyi, *Res. Technol. Mgmt.*, 45–55 (May–June 1994).
71. J. D. Holmes, "Doubling Research's Output Using Total Quality Management," *Industrial Research Institute Meeting*, Williamsburg, Va., Oct. 23–26, 1994.
72. J. D. Holmes and D. J. McCloskey, in *1992 American Society for Quality Control, Quality Congress Transactions*, American Society for Quality Control, Toronto, Canada, 1992, p. 531.
73. J. D. Holmes, G. O. Nelson, and D. C. Stump, *Res. Technol. Mgmt.*, 27–35 (May–June 1993).

74. J. D. Holmes and G. E. McGraw, *Res. Technol. Mgmt.*, 20–24 (Sept.–Oct. 1994).

75. M. E. McGrath, M. T. Anthony, and A. R. Shapiro, *Product Development: Success Through Product and Cycle Time Excellence*, Butterworth-Heinemann, Newton, Mass., 1992.

76. P. F. Drucker, *Harvard Bus. Rev.*, 69–79 (Nov.–Dec. 1991).

77. R. Szakonyi, ed., *Technology Management-2*, Auerbach Publications, Boston, Mass., 1993.

78. R. Szakonyi, ed., *Technology Management-3*, Auerbach Publications, Boston, Mass., 1994.

79. R. Szakonyi, ed., *Technology Management*, Auerbach Publications, Boston, Mass., 1995.

80. M. L. Tushman and Wil Moore, *Readings in the Management of Innovation*, 2nd ed., Harper Business, Boston, Mass., 1988.

81. *The Effective Management of Technical Functions in the Process Industries*, SRI International, Menlo Park, Calif., 1993.

82. J. M. Utterbach, *Mastering the Dynamics of Innovation*, Harvard Business School Press, Boston, Mass., 1994.

83. *R&D Productivity*, Hughes Aircraft Co., Culver City, Calif., 1978.

84. M. L. Patterson, *Accelerating Innovation*, Van Nostrand Reinhold, New York, 1993.

85. E. B. Roberts, *Res. Technol. Mgmt.*, 11–29 (Jan.–Feb. 1988).

86. A. C. Perrino and J. W. Tipping, *Res. Technol. Mgmt.*, 12–19 (May–June 1989).

87. R. E. Burkart, *Res. Technol. Mgmt.*, 27–32 (May–June 1994).

88. R. Szakonyi, *Res. Technol. Mgmt.*, 31–36 (July–Aug. 1990).

89. R. Szakonyi, *Res. Technol. Mgmt.*, 41–46 (Nov.–Dec. 1990).

90. P. F. Drucker, *Wall St. J.* **A22**, 90 (May 30, 1989).

91. W. E. Souder, *Managing New Product Innovations*, Lexington Books, Lexington, Mass., 1987, pp. 239–244.

92. R. N. Foster, L. H. Linden, R. L. Whiteley, and A. M. Kantrow, *Res. Technol. Mgmt.*, 17–26 (Jan.–Feb. 1985).

93. R. N. Foster, L. H. Linden, R. L. Whiteley, and A. M. Kantrow, *Res. Technol. Mgmt.*, 13–22 (Mar.–Apr. 1985).

94. P. M. Norling, "Managing Technology: Critical Questions for the Synthetic Rubber Industry," *29th Annual Meeting, International Institute of Synthetic Rubber Producers*, Quebec City, Canada, May 1988.

95. D. Matheson, J. E. Matheson, and M. M. Menke, *R&D Decision Quality Association Benchmarking Study*, Strategic Decisions Group, Menlo Park, Calif., 1993.

96. D. L. Ransley and J. L. Rogers, *Res. Technol. Mgmt.*, 19–26 (Mar.–Apr. 1994).

97. P. A. Schumann, Jr., D. L. Ransley, and D. C. L. Prestwood, *Res. Technol. Mgmt.*, 45–54 (May–June 1995).

98. L. Lander, D. Matheson, M. M. Menke, and D. L. Ransley, *Res. Technol. Mgmt.*, 40–43 (Jan.–Feb. 1995).

99. J. A. Trainham, personal communication, DuPont Co., Nov. 15, 1994.

100. M. A. Cusumano and R. W. Selby, *Res. Technol. Mgmt.*, 26–30 (Jan.–Feb. 1996).

101. G. A. Stevens and J. Burley, *Res. Technol. Mgmt.* (1996) in press.

102. J. A. Miller, *Res. Technol. Mgmt.*, 9–10 (Sept.–Oct. 1995).

103. G. Bylinsky, *Fortune*, 80A–80J (Jan. 15, 1996).

104. P. Gwynne, *Res. Technol. Mgmt.*, 30–33 (Jan.–Feb. 1995).

105. H. W. Chesbrough and D. J. Teece, *Harvard Bus. Rev.*, 65–73 (Jan.–Feb. 1996).

106. F. Kodama, *Harvard Bus. Rev.*, 70–78 (July–Aug. 1992).

107. C. W. Prather and L. K. Gundry, *Blueprints For Innovation: How Creative Processes Can Make You And Your Company More Competitive*, American Management Association, New York, 1995.

J. A. Miller
P. M. Norling
J. W. Collette
DuPont Central Research and Development

RESINS, NATURAL

Natural resins are generally described as solid or semisolid amorphous, fusible, organic substances that are formed in plant secretions. They are usually transparent or translucent yellow-to-brown colored, and are soluble in organic solvents but not in water. The principal uses for natural resins are in varnishes, printing inks, adhesives, paper size, and polymer compositions. The term natural resins includes tree and plant exudates, fossil resins, mined resins, and shellac. They often have been altered from their original state during isolation and processing. For some applications, the resins have been chemically modified to increase their industrial utility.

Natural resins, except for shellac, are mixed condensation products of naturally occurring terpenoids and flavonoids contained in trees. Shellac is a product of insect secretion. Molecular weights probably do not exceed 10,000. The resins do not exhibit simple melting or boiling points, although they do have a predominant glass-transition temperature, T_g . The T_g of natural resins ranges from about 0 to 100°C. The relatively high T_g is caused by the rigid, cyclic structure of the resin. Depending on its T_g , a natural resin at room temperature may be a brittle glass, or it may be flexible and extensible.

Natural resins were probably known to early people, who recognized them as exudates from trees. Collection and use of these resins have been recorded by early Roman and Greek historians. Many products have been collected by the same methods throughout history to the present time. However, increased labor costs and competition from synthetic resins have reduced the demand for some natural resins, so they have become less available. In other cases, such as that of rosin, the traditional collection of gum from trees has been supplemented or replaced by isolation from other sources, such as paper pulping and tree stumps.

Rosin and Modified Rosins

Rosin and associated products obtained from pine trees have been used for hundreds of years to caulk the bottoms of wooden sailing vessels and to lubricate the lines. These materials are known as naval stores because of this practice.

Production. Rosin is isolated from pine trees, principally from longleaf *Pinus palustris*, slash *Pinus ellioti*, and loblolly pine *Pinus taeda*. The products are known as gum, wood, or tall oil rosin, based on the method of isolation and the source.

In the gum rosin process, pine trees are wounded to stimulate the flow of gum. V-shaped slashes are cut through the bark, and the exudate is collected in a bucket below the slash. Production is stimulated by painting sulfuric acid on the slashes. The oleoresin (exudate) is separated by distillation into gum spirits of turpentine and gum rosin. The gum turpentine industry has decreased in importance in the 1990s because it is labor-intensive. The process is carried out in Russia, the People's Republic of China, Indonesia, Portugal, Brazil, and Mexico.

In the wood rosin process, rosin is isolated from aged pine stumps that have been left in fields cleared for farming or lumbering operations. The stumps are cut and shredded to pieces the size of matchsticks. The wood chips are then extracted with an appropriate solvent, eg, aliphatic or aromatic petroleum hydrocarbons or ketones. The extract is fractionally separated into nonvolatile crude rosin, volatile extractibles, and recovered solvent. The dark rosin is usually refined further to lighter-colored products using selective solvents or absorption.

Tall oil rosin is a by-product of paper manufacturing. Raw wood chips are digested under heat and pressure with a mixture of sodium hydroxide and sodium sulfide. Soluble sodium salts of lignin, rosin, and fatty acids are formed, which are removed from the wood pulp as a dark solution. The soaps of the rosin and fatty acids float to the top of the mixture, where they are skimmed off and treated with sulfuric acid to free the rosin and fatty acids. This mixture, known as crude tall oil (CTO), is refined further to remove color and odor bodies; fractional distillation separates the tall oil rosin acids from the fatty acids (see TALL OIL).

The three types of rosin perform differently because of different distributions of resin acids, as well as different types and quantities of impurities. These differences are reflected in end-use performance.

Properties. Rosin is a brittle, friable solid which has a T_g of ca 30°C and a ring and ball (R&B) softening point of ca 75°C. Although the average molecular weight is only about 300, its rigid cyclic structure results in a higher T_g than would be expected for an amorphous polymer of the same molecular weight. Heating above the softening temperature lowers viscosity sharply as a consequence of the low molecular weight.

Rosin is compatible with many materials because of its polar functionality, cycloaliphatic structure, and its low molecular weight. It has an acid number of ca 165 and a saponification number of ca 170. It is soluble in aliphatic, aromatic, and chlorinated hydrocarbons, as well as esters and ethers. Because of its solubility and compatibility characteristics, it is useful for modifying the properties of many polymers.

Composition. Rosin is primarily a complex mixture of monocarboxylic acids of alkylated hydrophenanthrene nuclei. These constituents, known as resin acids, represent about 90% of rosin. The resin acids are subdivided into two types, based on their skeletal structure. The abietic-type acids contain an isopropyl group pendent from the carbon numbered 13. The pimaric-type acids have a methyl and vinyl group pendent from the same carbon atom. Figure 1 shows the structure of typical resin acids; abietic acid, C₂₀H₃₀O₂ (**1**) is predominant. The remaining 10% of commercial rosin consists of neutral materials that are either hydrocarbons or saponifiable esters. These materials are derived from resin acids by decarboxylation or esterification. Typical grades of tall oil rosin also contain about 2–5% of C₁₈–C₂₂ fatty acids.

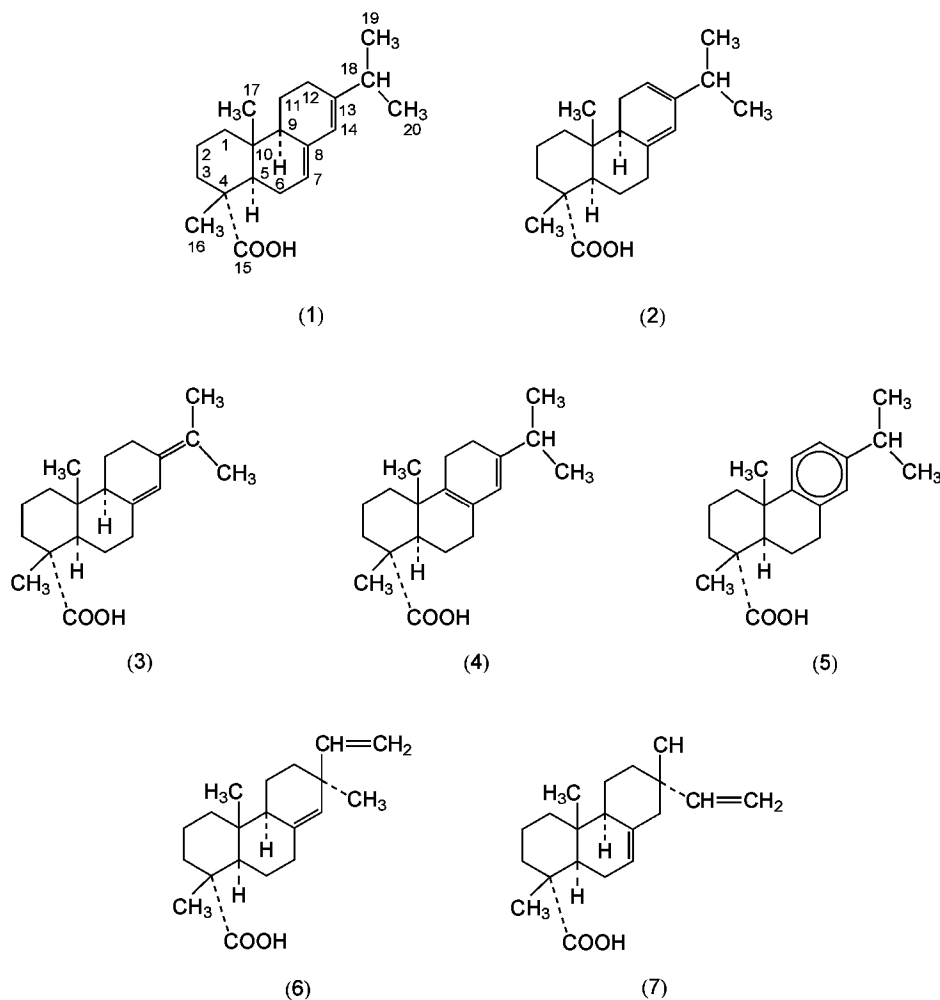
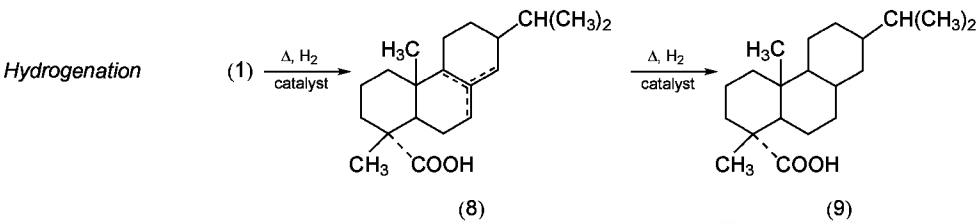


Fig. 1. Abietic-type acids include abietic acid [514-10-3] (1), levopimaric acid [79-54-9] (2), neoabietic acid [471-77-2] (3), palustric acid [1945-53-5] (4), and dehydroabietic acid [1740-19-8] (5). Pimonic-type acids are pimonic acid [127-27-5] (6) and isopimonic acid [5835-26-7] (7).

Modified Rosins. Natural rosins are often modified to improve their stability to long-term aging and oxidative degradation. Pale color and color stability are also important. Stabilization methods involve the double bonds in the resin acid molecule that can form a conjugated system highly susceptible to degradation. The double bonds can be stabilized by hydrogenation, disproportionation, or polymerization. Hydrogenation stabilizes the rosin by eliminating

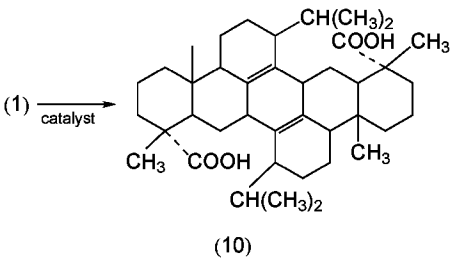


at least one of the double bonds; dihydroabietic acids (8) and tetrahydroabietic acid (9) are formed. Hydrogen also reacts with color bodies to stabilize the product further.

In disproportionation, rosin is heated over a catalyst to transfer hydrogen, yielding dehydro (5) and dihydro (8) resin acids. The dehydro acids are stabilized by the aromatic ring; the dihydro acids contain only an isolated double bond in place of the less stable conjugated double bonds.

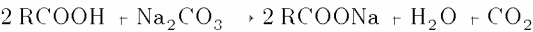


Rosin is polymerized by acid treatment at elevated temperatures, whereby double bonds are eliminated by dimerization. Dimerized rosin (10) has a higher molecular weight and a correspondingly higher softening point. It is more stable because fewer double bonds remain and conjugation is reduced.



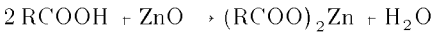
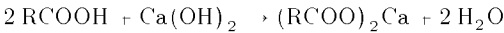
Derivatives. Most rosin products are sold as derivatives, which extends their utility. The carboxyl group or the double bonds are usually involved in the reactions.

The carboxyl group reacts with metal oxides, hydroxides, or salts to form rosin soaps or salts (resinates). The soaps of alkali metals, such as sodium and potassium, are useful in paper sizing and as emulsifiers in rubber polymerization.



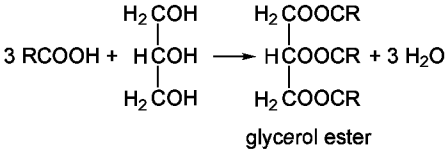


The salts, primarily calcium and zinc resinate, are used in printing ink formulations. The carboxyl group is converted to the alcohol by catalytic

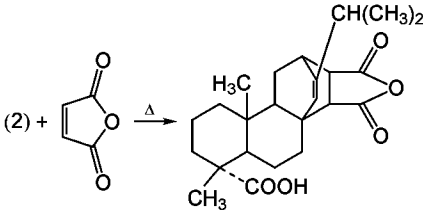


hydrogenation. Primary amines are prepared by amination to the amide, further dehydration to the nitrile, followed by hydrogenation to the amine.

The most important derivatives of the carboxyl group are formed by esterification with monohydric or polyhydric alcohols. Typical alcohols used include methyl alcohol, ethylene glycol, glycerol, and pentaerythritol. These rosin esters have a wide range of softening points and compatibilities.



The Diels-Alder reaction with dienophiles such as maleic anhydride or acrylic acid gives useful polycarboxylic acids, which can then be converted to metal salts or soaps or esterified to esters with high softening points.



Uses. The largest commercial use of rosin is in sizing paper to improve its water resistance, but it is being displaced by synthetic sizes. Paper size consists of rosin or modified rosins that have been partially saponified with sodium carbonate or sodium hydroxide. Sizes prepared from rosin may be modified with maleic anhydride to increase efficacy. Specialty paper is treated with stabilized rosins to minimize yellowing during aging. Sizes are used in concentrations of 0.1–3.0%. Aluminum sulfate (alum) (0.5–3.0%) increases the acidity of the paper pulp slurry. The sodium soaps of the rosin or modified rosins are precipitated as insoluble, positively charged aluminum mono- and diresinates, which are attracted to the negatively charged pulp fibers. They are retained in the paper mat when the water is removed. The hydrophobic aluminum resinate resists the penetration of water into the capillaries of the paper. Sizes consisting of 30–50% aqueous dispersions of rosin are also used. The alum–rosin interaction occurs on the surface of the pulp fiber.

Modified rosins, metal resinates, and resin acid esters find application in printing ink formulations. Rosin-derived resins dispersed in linseed oil are used as vehicles for letterpress inks. The resins provide specific adhesion, abrasion resistance, and gloss, and they improve pigment wetting and dispersion. Resins with high acid numbers are used as vehicles for flexographic inks, which are designed for printing on paper, metal, and plastic films. The resins provide flexibility, abrasion resistance, and gloss. Metal salts of rosins, modified rosins, and polymerized rosins are used in inexpensive gravure inks. In higher quality inks, rosin ester resins and phenolic-modified rosin esters provide hardness, gloss, and adhesion.

Rosin ester resins are used extensively in pressure-sensitive adhesives as tackifiers. The adhesive is formulated by blending the resin with a polymer in solution or as aqueous emulsions. Typical compositions may contain about 50% rosin. The glycerol or pentaerythritol esters of stabilized rosins are often used because they are stable on aging.

Rosin, modified rosins, and derivatives are used in hot-melt adhesives. They are based primarily on ethylene–vinyl acetate copolymers. The rosin derivative is used in approximately a 1:1:1 concentration with the polymer and a wax. The resin provides specific adhesion to the substrates and reduces the viscosity at elevated temperatures, allowing the adhesive to be applied as a molten material.

Rosin ester resins are used as modifiers in the formulation of chewing gum. The rosin derivative modifies the physical properties of the polymer used, providing the desired masticatory properties. The glycerol ester of hydrogenated rosin is the predominant choice, because stabilized materials have improved aging resistance, which extends the shelf life of the gum.

The soap of modified rosin has a long history as an emulsifier for the polymerization of styrene–butadiene rubber. The rosin soap remains in the rubber after polymerization and increases the tack between the plies of rubber required in the manufacture of tires.

Traditional Natural Resins

Natural resins have been collected by hand throughout recorded history and used with minimal processing. They are reported to have been used in the arts, both in paints and for polishing sculptures, as early as 350 BC. Amber, the hardest of these resins, has been used as a gemstone from early Greek history to modern times. The electrical properties of amber were first recorded about 300 BC. Following is a description of commercial natural resins that are available in the United States.

Manila Copal. The Manilas are collected in Indonesia and the Philippines. They are soluble in alcohols and ketones, and insoluble in hydrocarbons and esters. The resins soften between 81–90°C and have acid numbers of 110–141. Principal uses are in coatings and varnishes.

Pontianak. This resin is a copal and is similar to the alcohol-soluble Manilas. It is partially fossilized, so it melts at a higher temperature. Softening points range from 99–135°C, and acid numbers from about 112–120. Pontianak [9000-14-0] is used in specialty coatings and adhesives.

Dammar Resins. These resins are tapped from trees in Indonesia and Malaysia. They have low acid numbers, typically 17–35, and softening points of between 67–75°C. They are soluble in aliphatic and aromatic hydrocarbons and in terpenes. Dammar resins [9000-16-2] are used primarily in protective-coating formulations. For use in lacquers, a small, high melting fraction of raw dammar must be removed to increase solubility in common lacquer solvents. The softening point of dewaxed dammar is 77–81°C, and its acid number ranges from 25–31. It is used as a spirit varnish over artistic oil paintings to shield against dirt.

East India Resins. The East India resins are related to the dammars, although they are older and harder. They are not obtained by tapping trees, but are collected where they are found, principally in Indonesia. Because they are semifossil resins, their softening points are high, ranging from about 110–130°C. The East India resins [9000-16-2] have low (20–30) acid numbers. They are soluble only in aryl hydrocarbons and hydrogenated aliphatic hydrocarbons, and are used primarily in coatings.

Gum Elemi. This resin, tapped from trees in the Philippines, contains a higher concentration of essential oils than other natural resins. It is a soft, sticky, plastic material that can be deformed manually. Gum elemi [9000-75-3] contains 20–25% essential oils, 13–19% acids, 30–35% resenes (condensed decarboxylated resin acids), and 20–25% terpenic resinols (condensed terpene alcohols). It has an acid number of 20–35 and a saponification number of 20–40. Gum elemi is a film-forming plasticizing resin used in lacquers.

Sandarac. This resin, which originates in Morocco, is a polar, acidic, hard resin with a softening point of 100–130°C, an acid number of 117–155, and a saponification number of 145–157. Sandarac [9000-57-1] is soluble in alcohols and insoluble in aryl and aliphatic hydrocarbons. It is used in varnishes and lacquers for coating paper, wood, and metal.

Mastic. Most commercial mastic [61789-92-2] is collected on the Greek island of Chios, near the Turkish coast. It is a soft resin with a softening point of 55°C. It has an acid number of 50–70 and a saponification number of 62–90. It is soluble in alcohols and aryl hydrocarbons. Mastic is used in wood coatings, lacquers, adhesives, and printing inks.

Other Resins. Some resins that were important in earlier years have decreased in commercial significance because of scarcity and increased cost of collection. They are listed for the historical record.

Kauri. This fossil resin, classified as a copal, is found in the South Pacific, primarily in New Zealand. It formerly was used in protective coatings. It is still used in the Kauri-butanol test (ASTM D1133) to determine the volume of thinner that can be added to a varnish formulation without causing turbidity.

Congo. Another fossil copal, congo was developed by Belgian interests in the early 1900s in what is now Zaire. Production diminished when the Belgians left, and in the 1990s there is no organized isolation of congo in Zaire.

Accroides. This resin is collected in Australia in small quantities. Accroides has been used as a binder for wallboard, and in coatings and printing inks.

Amber. The hardest of all the resins, amber is still collected along the shores of the Baltic Sea. The pieces are polished to an attractive pale yellow-to-dark brown stone and made into ornamental objects.

Natural Resins in Medicines, Flavors, and Fragrances

Some natural resins have been used for many centuries in medical applications, flavors, and fragrances. With advances in pharmacology, the medicinal uses of these resins have mostly disappeared. The only remaining significant application is the use of guaiac-impregnated paper to detect hemoglobin in stool. This widely used test is helpful in the early diagnosis of colorectal cancer. Use of the resins in cosmetics (qv) has come into question because of reports of contact dermatitis and skin sensitivity.

The natural resins described here are those that have been mentioned in technical literature; they can be identified by CAS Registry Numbers and Merck Index numbers (1).

Balm of Gilead [8022-26-2], also known as Balsam Mecca (Merck No. 958), obtained from Arabia and Abyssinia, has been used in perfumery. Balsam of Peru [8007-00-9] (Merck No. 959), obtained from Salvador, has been used in perfumery, skin ulcer therapy, and as an ingredient in chocolate flavoring. Balsam of Tolu [9000-64-0] (Merck No. 960), obtained from Venezuela and Colombia, has been used in perfumery, chewing gums, and as an expectorant. Copaiba [8001-61-4] (Merck No. 2511), obtained from the Caribbean and Brazil, has been used for treatment of cystitis, gonorrhea, and bronchial diseases, and as a diuretic. It has also been used as a varnish remover for oil paintings. Gum benzoin [9000-05-9] (Merck No. 4492), obtained from Indonesia and Thailand, has been used in ointments as a topical protector. Guaiac [9000-29-7] (Merck No. 4455), obtained from the Caribbean and the northern coast of South America, is still actively used as a reagent in testing for blood in stool. Myrrh [9000-45-7] (Merck No. 6251), obtained from East Africa and Arabia, has been used as a carminative and as an astringent for mucous membranes. Its ancient use as a perfume and as an incense is first described in the Bible. Olibanum [8050-07-5] (Merck No. 6792), also known as frankincense, is obtained from Arabia. It has been used as an ingredient of incense. Storax [8046-19-3] (Merck No. 8778), obtained from Turkey, has been used as a topical protector and an expectorant.

Asphaltites

Mining operations are required to bring asphaltites to the market, principally for use in coatings and inks. They were formed in the same manner as oil and coal, below the surface of the earth from vegetation that had been subjected to high temperatures and pressures on a geological time scale. The most widely known asphaltite is gilsonite, which is found in the Uinta Basin in eastern Utah. Gilsonite has a softening point of 110–121°C. It has been used in protective coatings, battery cases, asphalt floor tiles, brake linings, and inks.

Glance pitch, or manjak, and grahamite are related to gilsonite, but have higher specific gravities. Glance pitch has the same softening point as gilsonite, 110–121°C. Grahamite is the highest melting of the asphaltites, softening at 176–315°C.

Shellac

Unlike other natural resins, shellac [9000-59-3] is derived from the hardened secretion of the lac insect (species *Laccifer (Tachardia) lacca* Kerr (family Coccidae), also known as *Kerris lacca* (Kerr)). Shellac is a refined grade of the crude lac secretion and is the most widely known lac product. Therefore, shellac has been accepted as the common generic term. Over 50% of the world's supply is produced in the Indian provinces of Bihar and Orissa, with the remainder originating in adjacent areas of southeast Asia such as Sri Lanka, China, Thailand, and Myanmar.

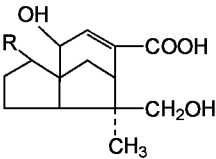
The collection and use of lac by Indian aboriginal tribes probably predates recorded history. Ancient Greek and Roman writers were aware of lac. By the late fifteenth century, European craftsmen were attracted to the use of lac as a finish for cabinets and other furniture because of its gloss and luster. Lac became an important component of decorative and protective finishes by the nineteenth century. It is ironic that the success of shellac led to the synthetic resin industry. Baekeland developed phenolic resins while trying to find a substitute for shellac.

Lac is produced by the larvae of the lac insect, which secretes a coating composed of lac to protect it from weather and enemies. The secretions of neighboring larvae coalesce into one continuous coating over twigs of the host tree. This coating is harvested by cutting the twigs from the trees or by twisting the branches so that the resin falls off.

Raw lac is first treated to remove water-soluble carbohydrates and the dye that gives lac its red color. Also removed are woody materials, insect bodies, and trash. It is further refined by either hot filtration or a solvent process. In the heat process, the dried, refined lac is filtered molten through cloth or wire screens to produce the standard grades of orange shellac. In the solvent process, lac is dissolved and refluxed in alcohol solvents, filtered to remove dirt and impurities, and concentrated by evaporation. The lac can be further decolorized in this process to produce very pale grades. Bleached shellac is prepared by treatment with dilute sodium hypochlorite and coalesced into slabs.

Composition. Shellac is primarily a mixture of aliphatic polyhydroxy acids in the form of lactones and esters. It has an acid number of ca 70, a saponification number of ca 230, a hydroxyl number of ca 260, and an iodine number of ca 15. Its average molecular weight is ca 1000. Shellac is a complex mixture, but some of its constituents have been identified. Aleuritic acid, an optically inactive 9,10,16-trihydroxypalmitic acid, has been isolated by saponification. Related carboxylic acids such as 16-hydroxy- and 9,10-dihydroxypalmitic acids, also have been identified after saponification. These acids may not be primary products of hydrolysis, but may have been produced by the treatment. Studies show that shellac contains carboxylic acids with long methylene chains, unsaturated esters, probably an aliphatic aldehyde, a saturated aliphatic ester, a primary alcohol, and isolated or unconjugated double bonds.

Cyclic sesquiterpenes having the cedrene skeleton have also been isolated. When R = COOH, this structure (11) has been named shellolic acid. Jalaric acid (11, R = CHO) and laksholic acid (11, R = CH₂OH) also have been isolated.



Uses. Synthetic resins have taken over a large share of the market for shellac. Unpigmented shellac is used on floors, woodwork, and paneling.

White pigmented shellac is used as a primer–sealer for interior applications. Shellac is used as a protective coating for pharmaceuticals to maintain the potency of medication. Several coats of shellac glaze protect medication from the effects of stomach acids so that the medication is not released before the tablet reaches the intestines.

Candy is coated with shellac to seal in moisture and keep the product fresh. The coating provides a high gloss to the confection, which improves its appearance. Citrus fruits and some apples are often coated with shellac. This improves the appearance, while allowing the fruit to breathe without spoilage. Shellac is used as a stiffener for felt hat bodies, primarily for recreational hats. It is also used to stiffen playing cards, providing "snap."

Economic Aspects

Rosin and its derivatives are economically the most important natural resins. Approximately 1150 × 10³ metric tons of these materials are produced annually and sold throughout the world. The principal producers are the People's Republic of China (ca 40^o %) and the United States (ca 25^o %), followed by Russia. Most of the remainder is produced in Indonesia, Portugal, Finland, India, Brazil, and Mexico. In 1996, the lowest grades of rosin were priced at \$750 t. Most rosin is converted to its many derivatives to meet requirements for industrial applications. The principal producers of rosin derivatives are Arizona Chemical Company, Hercules Incorporated, Westvaco, Union Camp, and Georgia-Pacific.

The traditional natural resins are collected or isolated from trees, primarily in the more moderate climates of the world. Before World War II, annual consumption of these resins in the United States was about 18,000–23,000 t. This dropped to about 9000 t yr by the late 1940s. The total imported volume in 1995 is estimated at <500 t. These resins have been replaced by synthetic resins in most industrial applications. Traditional natural resins are sold in bulk quantities for about \$1.32–\$6.60 kg. Special grades of these resins are sold for as much as \$132 kg. The largest importer of traditional natural resins is P. L. Thomas & Company, Inc.

Shellac has had a commercial history similar to that of the traditional resins. It has been replaced by synthetic resins in many applications. In the 1950s, about 19,000 t of various grades of processed shellac were available in the United States. The volume was about 5,500 t in 1995. The largest importer of shellac is William Zinsser & Company, Inc. The superior grades of shellac are sold for about \$6.60–\$9.90 kg, depending on quality.

Health and Safety Information

Rosin has a low order of toxicity following ingestion or skin contact. Rosin and its numerous derivatives have a number of permitted food packaging and other direct and indirect food contact uses throughout the world. Sanctioned uses applicable in the United States are outlined in U.S. Food and Drug Administration (U.S. FDA) Regulations (2). Material Safety Data Sheets (MSDSs) for specific rosins and their derivatives should be consulted before their use.

Natural resins such as dammar and Manila copal have been described in U.S. FDA Regulations (3). The Material Safety Data Sheets for these products issued by the importer describe them as nontoxic and nonallergenic.

Shellac has been affirmed by the U.S. FDA (4) as GRAS for food applications. The most significant uses for shellac are in food, for coating fruits and confections to preserve quality and appearance.

BIBLIOGRAPHY

"Resins, Natural" in *ECT* 1st ed., Vol. 11, pp. 666–687, by E. L. Kropa, Battelle Memorial Institute; in *ECT* 2nd ed., Vol. 17, pp. 379–391, by F. A. Wagner, B. F. Goodrich Chemical Co.; in *ECT* 3rd ed., Vol. 20, pp. 197–206, by J. C. Weaver, Case Western Reserve University; "Rosin and Rosin Derivatives" in *ECT* 3rd ed., Vol. 17, pp. 475–508, by H. I. Enos, Jr., and G. C. Harris, Hercules Incorporated, and G. W. Hedrick, Naval Stores Laboratory, U.S. Dept. of Agriculture; "Shellac" in *ECT* 3rd ed., Vol. 20, pp. 737–747, by J. Martin, William Zinsser & Co.

1. S. Budavari, M. J. O'Neil, A. Smith, and P. E. Heckelman, *The Merck Index*, 11th ed., Merck and Co., Inc., Rahway, N.J., 1989.
2. "Rosin and Rosin Derivatives," *Code of Federal Regulations*, Title 21, Part 178.3870, U.S. Food and Drug Administration, Washington, D.C., Apr. 1995.
3. "Resinous and Polymeric Coatings," in Ref. 2, Part 175.300, Apr. 1995.
4. Ref. 2, Part 184.1705, 1989.

General References

J. B. Class, "Resins, Natural," in J. I. Kroschwitz, ed., *Encyclopedia of Polymer Science and Engineering*, 2nd ed., Vol. 14, John Wiley & Sons, Inc., New York, 1988, pp. 438–453.

Rosin and Modified Rosin

N. Heaton, *Outlines of Paint Technology*, Charles Griffin & Co., London, 1947, pp. 280–305. Also contains information on classical natural resins and shellac. Technical literature, Hercules Incorporated, Wilmington, Del. Information also available on mined resins.

Traditional Natural Resins

C. L. Mantell, C. W. Kopf, J. L. Curtis, and E. M. Rogers, *The Technology of Natural Resins*, John Wiley & Sons, Inc., New York, 1942. *Natural Resins Handbook*, American Gum Importers Association, New York, 1939. G. T. Sohl, in W. Von Fischer, ed., *Paint and Varnish Technology*, Reinhold Publishing Corp., New York, 1948, pp. 119–126. R. O. Innes, *Mod. Paint Coat.* **72**(9), 51–52 (1982). Technical literature, O. G. Innes Corp., New York.

Resins in Medicine, Flavors, and Fragrances

S. Budavari, M. J. O'Neil, A. Smith, and P. E. Heckelman, *The Merck Index*, 12th ed., Merck and Co., Inc., Rahway, N.J., 1995. H. M. Boelens, D. deRijke, and H. G. Haring, *Perfumer Flavor.* **6**(6), 7–14 (1981). B. M. Lawrence, *Perfumer Flavor.* **7**(1), 45–50 (1982).

Mined Resins

H. Abraham, *Asphalts and Allied Substances*, 5th ed., D. Van Nostrand Co., New York, 1945, p. 220.

Shellac

J. W. Martin, in R. R. Meyers and J. S. Long, eds., *Treatise on Coatings*, Vol. I, *Film-Forming Compositions*, Part II, Marcel Dekker, Inc., New York, 1968, pp. 441–478. W. H. Gardner, in J. J. Mattiello, ed., *Protective and Decorative Coatings*, John Wiley & Sons, Inc., New York, 1941, pp. 259–291. Technical literature, William Zinsser & Co., Inc., Somerset, N.J.

Jay B. Class
Hercules Incorporated

RESINS, WATER-SOLUBLE.

See WATER SOLUBLE POLYMERS.

RETROVIRUSES.

See ANTIVIRAL AGENTS; IMMUNOTHERAPEUTIC AGENTS.

RESORCINOL.

See HYDROQUINONE, RESORCINOL, AND CATECHOL.

REVERSE OSMOSIS

Reverse osmosis (RO) is a fairly mature technology used in the area of seawater and brackish water desalination (see WATER, SUPPLY AND DESALINATION). Use of reverse osmosis as a separation tool is, however, a relatively young technology. In the late 1950s, it was shown that cellulose acetate RO membranes were capable of separating salt from water, although the water fluxes obtained were too small to be practical (1–4). Then in the early 1960s, a method for making asymmetric cellulose acetate membranes having relatively high water fluxes and separations was developed, making RO separations both possible and practical (5–7).

Since that time, the industrial development of thin-film composite membranes (8) (see MEMBRANE TECHNOLOGY) and improvements in polymer materials have widened applications of RO from desalination to treatment of hazardous wastes (see WASTE TREATMENT, HAZARDOUS WASTE), water purification uses, and material recovery in the electroplating (qv) industries. These membranes can tolerate wide pH ranges, higher temperatures than cellulose acetate, and harsh chemical environments. Moreover, water flux and solute separation characteristics have highly improved, resulting in many applications. For example, these RO membranes have found uses in wastewater treatment, production of ultrapure water, water softening, and food processing (qv) (9). Sales of RO membrane products were as high as \$118 million annually in the late 1980s, and continued growth is expected (10).

The driving force for the development and use of RO membranes is the advantage of these systems over traditional separation processes such as distillation (qv), extraction, ion exchange (qv), and adsorption (qv). Because reverse osmosis is a pressure-driven process, no energy-intensive phase changes or potentially expensive solvents or adsorbents are needed for RO separations. Reverse osmosis is a process that is inherently simple to design and operate compared to many traditional separation processes. Also, simultaneous separation and concentration of both inorganic and organic compounds are possible using the RO process. In addition, using nanofiltration, ie, loose RO, membrane-selective solute separations based on charge and molecular weight and size differences are possible. Finally, reverse osmosis technology can also be combined with ultrafiltration (qv), pervaporation, distillation, and other separation techniques to produce hybrid processes that result in highly efficient and selective separations. Excellent reviews of RO technology and theories have been reported in the literature (7,9–15).

Membrane Materials and Modules

Reverse osmosis membrane separations are governed by the properties of the membrane used in the process. These properties depend on the chemical nature of the membrane material, which is almost always a polymer, as well as its physical structure. Properties for the ideal RO membrane include low cost, resistance to chemical and microbial attack, mechanical and structural stability over long operating periods and wide temperature ranges, and the desired separation characteristics for each particular system. However, few membranes satisfy all these criteria and so compromises must be made to select the best RO membrane available for each application. Excellent discussions of RO membrane materials, preparation methods, and structures are available (8,13,16–21).

Most commercially available RO membranes fall into one of two categories: asymmetric membranes containing one polymer, or thin-film composite membranes consisting of two or more polymer layers. Asymmetric RO membranes have a thin (~100 nm) permselective skin layer supported on a more porous sublayer of the same polymer. The dense skin layer determines the fluxes and selectivities of these membranes whereas the porous sublayer serves only as a mechanical support for the skin layer and has little effect on the membrane separation properties. Asymmetric membranes are most commonly formed by a phase inversion (polymer precipitation) process (16). In this process, a polymer solution is precipitated into a polymer-rich solid phase that forms the membrane and a polymer-poor liquid phase that forms the membrane pores or void spaces.

An excellent review of composite RO and nanofiltration (NF) membranes is available (8). These thin-film, composite membranes consist of a thin polymer barrier layer formed on one or more porous support layers, which is almost always a different polymer from the surface layer. The surface layer determines the flux and separation characteristics of the membrane. The porous backing serves only as a support for the barrier layer and so has almost no effect on membrane transport properties. The barrier layer is extremely thin, thus allowing high water fluxes. The most important thin-film composite membranes are made by interfacial polymerization, a process in which a highly porous membrane, usually polysulfone, is coated with an aqueous solution of a polymer or monomer and then reacts with a cross-linking agent in a water-immiscible solvent.

Although RO membranes have been formed and tested using a wide range of different materials and preparation techniques, cellulosic polymers such as cellulose acetate and cellulose triacetate, linear and cross-linked aromatic polyamide, and aryl-alkyl polyetherurea are among the most important RO membrane materials (8,10,20). Asymmetric cellulose acetate membranes continue to enjoy widespread use despite some disadvantages, for example, a narrow (4.5–7.5) pH operating range, because the polymer is subject to hydrolysis, is susceptible to biological attack, exhibits compaction (mechanical compression) at high pressures that results in reduced water flux, and has low (~35°C) upper temperature limits. Polyamide and polyurea composite membranes typically have higher water fluxes and salt and organic rejections, an ability to withstand higher temperature and larger (4–11) pH variations, and immunity to biological attack and compaction. However, these membranes tend to be less chlorine-resistant and more susceptible to oxidation as compared to cellulose acetate membranes. Figure 1 shows water flux and NaCl rejections for three different classes of RO membranes made from a variety of polymer materials. Selected solute rejections, both inorganic and organic, for a large number of RO membranes can be found in the literature (9).

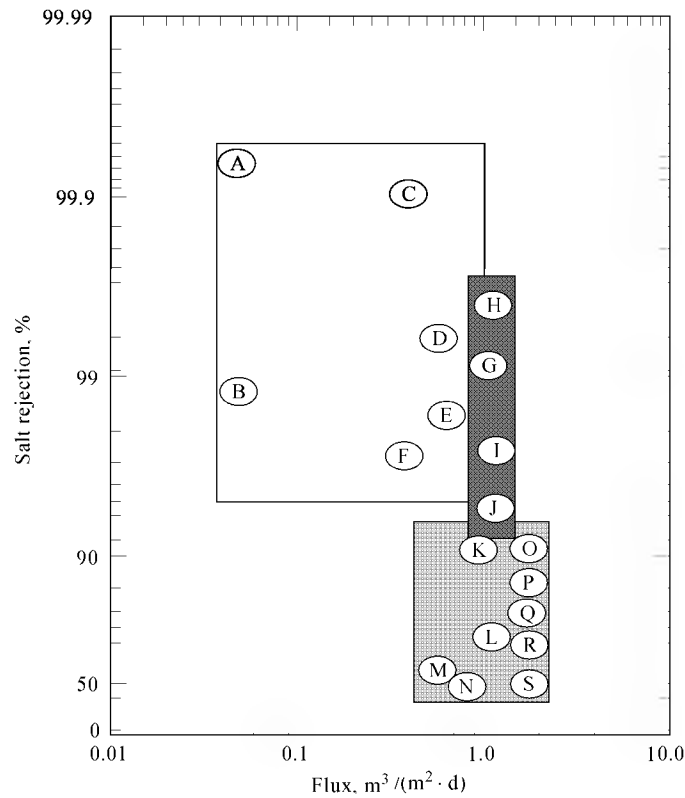


Fig. 1. Water flux and NaCl rejection of several membrane types (10), where (□) represents seawater membranes, which operate at 5.5 MPa and 25°C; (■), brackish water membranes, which operate at 1500 mg L NaCl feed, 1.5 MPa, and 25°C; and (▨) nanofiltration membranes, which operate at 500 mg L NaCl feed, 0.74 MPa, and 25°C. A represents cellulose acetate–cellulose triacetate; B, linear aromatic polyamide; C, cross-linked polyether; D, cross-linked fully aromatic polyamide; E, other thin-film composite membranes; F, asymmetric membranes; G, BW-30 (FilmTec); H, SU-700 (Toray); I, A-15 (Du Pont); J, NTR-739HF (Nitto-Denko); K, NTR-729HF (Nitto-Denko); L, NTR-7250 (Nitto-Denko); M, NF40 (FilmTec); N, NF40HF (FilmTec); O, UTC-40HF (Toray); P, NF70 (FilmTec); Q, UTC-60 (Toray); R, UTC-20HF (Toray); and S, NF50 (FilmTec). To convert MPa to psi, multiply by 145.

Whereas the membrane material largely determines the water and solute fluxes in an RO process, the packaging of the RO membrane is also important for process feasibility. Requirements of a membrane module include the offering of mechanical support to the RO membrane even at high (up to 8 MPa (1160 psi)) operating pressure, a design that minimizes pressure drop across the module as well as fouling and concentration polarization, and a module that is relatively inexpensive and easy to replace in the membrane process. The most common commercially available membrane modules include plate-and-frame, tubular, spiral-wound, and hollow-fiber elements (11). For large-scale desalination applications, spiral-wound modules have been commonly used. Another newer type of membrane module, Disc Tube (Rochem), which has reduced fouling problems, has been applied to various environmental problems ranging in sizes from 10 to 2000 m³ d (22,23). This system has also been reported to withstand operating pressures up to 27.6 MPa (4000 psi).

Theoretical Aspects

A reverse osmosis membrane acts as the semipermeable barrier to flow in the RO process, allowing selective passage of a particular species, usually water, while partially or completely retaining other species, ie, solutes such as salts. Chemical potential gradients across the membrane provide the driving forces for solute and solvent transport across the membrane. The solute chemical potential gradient, Δμ_s, is usually expressed in terms of concentration; the water (solvent) chemical potential gradient, Δμ_w, is usually expressed in terms of pressure difference across the membrane.

Measurable Process Parameters. The RO process is relatively simple in design. It consists of a feed water source, feed pretreatment, high pressure pump, RO membrane modules, and in some cases, post-treatment steps. A schematic of the RO process is shown in Figure 2a.

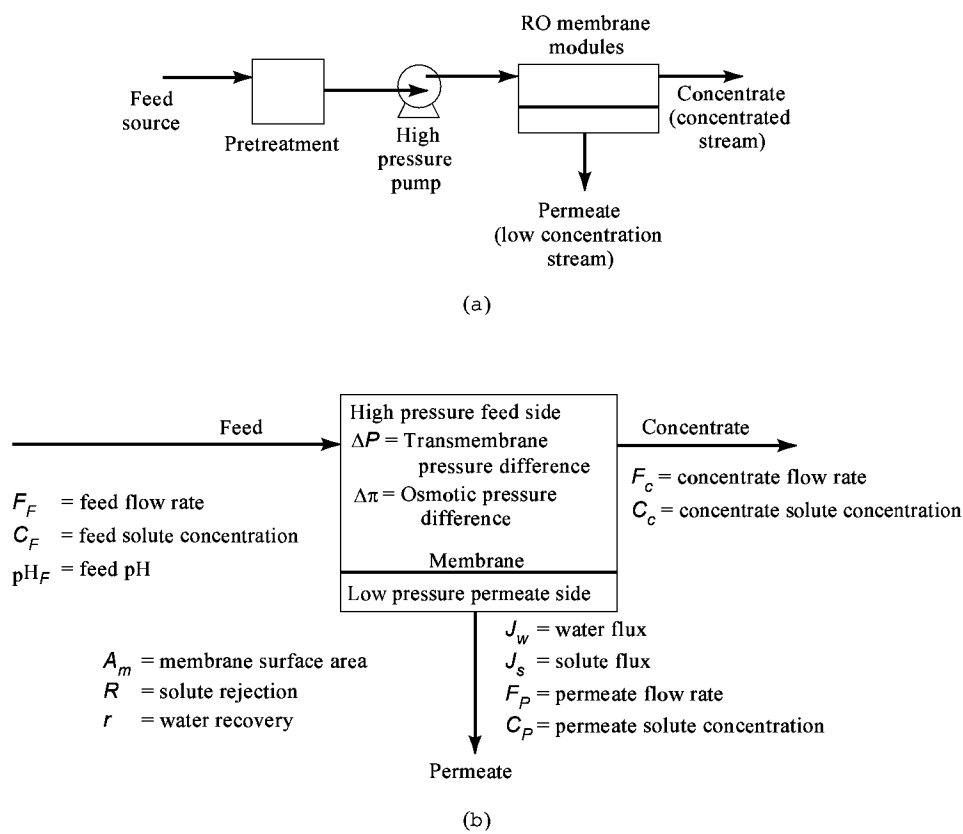


Fig. 2. Schematic of (a) an RO membrane process, and (b) the RO process streams.

The three streams and associated variables of the RO membrane process are shown in Figure 2b: the feed; the product stream, called the permeate; and the concentrated reject stream, called the concentrate or retentate. The water flow through the membrane is reported in terms of water flux, J_w .

$$J_w = \frac{\text{volumetric or mass permeation rate}}{\text{membrane area}}$$

Solute passage is defined in terms of solute flux, J_s .

$$J_s = \frac{\text{mass permeation rate}}{\text{membrane area}}$$

Solute separation is measured in terms of observed rejection, R , defined as

$$R = 1 - \frac{C_P}{C_F} \tag{1}$$

where C_P and C_F are the permeate and feed solute concentration, respectively. The quantity of feed water that passes through the membrane, ie, the permeate, is measured in terms of water recovery, r , defined for a batch RO system as

$$r = \frac{\Sigma J_w A_m \Delta t}{V_F} = \frac{V_P}{V_F} \tag{2}$$

where A_m is the membrane surface area; Δt the time interval, and V_F the volume of the feed; and for a continuous system as

$$r = \frac{J_w A_m}{F_F} = \frac{F_P}{F_F} \tag{3}$$

where F_F and F_P are the feed and permeate flow rates, respectively. In a batch membrane system, water is recovered from the system as the concentrate (retentate) is recycled to the feed tank. As a result, if the solute is rejected, the feed concentration, C_F , continuously increases over time. For a continuous membrane system, fresh feed is continuously supplied to the membrane.

Water flux is sometimes normalized according to the initial or pure water flux, J_{w0} , as J_w/J_{w0} , or as flux drop, defined by

$$\text{flux drop} = 1 - \frac{J_w}{J_{w0}} \tag{4}$$

The pressure difference between the high and low pressure sides of the membrane is denoted as ΔP ; the osmotic pressure difference across the membrane is defined as $\Delta \pi$; the net driving force for water transport across the membrane is $\Delta P - \sigma \Delta \pi$, where σ is the Staverman reflection coefficient and $\sigma = 1$ means 100% solute rejection. The standardized terminology recommended for use to describe pressure-driven membrane processes, including that for reverse osmosis, has been reviewed (24).

Transport Models. Many mechanistic and mathematical models have been proposed to describe reverse osmosis membranes. Some of these descriptions rely on relatively simple concepts; others are far more complex and require sophisticated solution techniques. Models that adequately describe the performance of RO membranes are important to the design of RO processes. Models that predict separation characteristics also minimize the number of experiments that must be performed to describe a particular system. Excellent reviews of membrane transport models and mechanisms are available (9,14,25–29).

Reverse osmosis models can be divided into three types: irreversible thermodynamics models, such as Kedem-Katchalsky and Spiegler-Kedem models; nonporous or homogeneous membrane models, such as the solution–diffusion (SD), solution–diffusion–imperfection, and extended solution–diffusion models; and pore models, such as the finely porous, preferential sorption–capillary flow, and surface force–pore flow models. Charged RO membrane theories can be used to describe nanofiltration membranes, which are often negatively charged. Models such as Donnan exclusion and the

extended Nernst-Planck include electrostatic effects. The transport models focus on the top thin skins of asymmetric membranes and of composite membranes because the skins determine fluxes and selectivities for most membranes. Also, most of the membrane models assume equilibrium, near-equilibrium, or steady-state conditions in the membrane.

A fundamental difference exists between the assumptions of the homogeneous and porous membrane models. For the homogeneous models, it is assumed that the membrane is nonporous, that is, transport takes place between the interstitial spaces of the polymer chains or polymer nodules, usually by diffusion. For the porous models, it is assumed that transport takes place through pores that run the length of the membrane barrier layer. As a result, transport can occur by both diffusion and convection through the pores. Whereas both conceptual models have had some success in predicting RO separations, the question of whether an RO membrane is truly homogeneous, ie, has no pores, or is porous, is still a point of debate. No available technique can definitively answer this question. Two models, one nonporous and diffusion-based, the other pore-based, are discussed herein.

Solution–Diffusion Model. In the solution–diffusion model, it is assumed that (1) the RO membrane has a homogeneous, nonporous surface layer; (2) both the solute and solvent dissolve in this layer and then each diffuses across it; (3) solute and solvent diffusion is uncoupled and each is the result of the particular material's chemical potential gradient across the membrane; and (4) the gradients are the result of concentration and pressure differences across the membrane (26,30). The driving force for water transport is primarily a result of the net transmembrane pressure difference and can be represented by equation 5:

$$J_w = \frac{D_{wm}C_{wm}\bar{V}_w}{R_gT\delta}(\Delta P - \Delta\pi) - A(\Delta P - \Delta\pi) \tag{5}$$

where D_{wm} and C_{wm} are the water membrane diffusivity and concentration, respectively; $\bar{V}_w$ is the partial molar volume of water; R_g , the gas constant; T , the temperature in Kelvin; δ , the membrane thickness; and A , the water permeability coefficient.

For the solute flux, it is assumed that chemical potential difference owing to pressure is negligible. Thus the driving force is almost entirely a result of concentration differences. The solute flux, J_s , is defined as in equation 6:

$$J_s = \frac{D_{sm}K_{sm}}{\delta}(C_{wall} - C_P) = B(C_{wall} - C_P) \tag{6}$$

where D_{sm} is the membrane solute diffusivity; K_{sm} , the solute partition coefficient for homogeneous membrane; C_{wall} and C_P , the feed solute concentration at the membrane wall and the permeate solute concentration, respectively; and B , the solute permeability coefficient. In the absence of concentration polarization, C_{wall} can be replaced by the bulk solute concentration, C_B or by the feed concentration, C_F for negligible water recovery. Using the relations for solvent and solute flux, solute rejection, R , for the solution–diffusion model can be expressed as in equation 7:

$$\frac{1}{R} = 1 + \left(\frac{B}{A}\right) \left(\frac{1}{\Delta P - \Delta\pi}\right) \tag{7}$$

Equation 7 shows that as $\Delta P \rightarrow \infty$, $R \rightarrow 1$. The principal advantage of the solution–diffusion (SD) model is that only two parameters are needed to characterize the membrane system. As a result, this model has been widely applied to both inorganic salt and organic solute systems. However, it has been indicated (26) that the SD model is limited to membranes having low water content. Also, for many RO membranes and solutes, particularly organics, the SD model does not adequately describe water or solute flux (27). Possible causes for these deviations include imperfections in the membrane barrier layer, pore flow (convection effects), and solute–solvent–membrane interactions.

When it was recognized (31) that the SD model does not explain the negative solute rejections found for some organics, the extended solution–diffusion model was formulated. The SD model does not take into account possible pressure dependence of the solute chemical potential which, although negligible for inorganic salt solutions, can be important for organic solutes (28,29).

Surface Force–Pore Flow Model. The surface force–pore flow (SFPF) model (8,32) is a two-dimensional extension of the finely porous model. Whereas the finely porous model considers only axial solute concentration gradients, the SFPF model recognizes that the solute concentration in an RO membrane pore may be a function of radial as well as axial position (29). The SFPF model assumes that (1) water transport through the membrane occurs in pores by viscous flow; (2) solute transport takes place by diffusion and convection in the membrane pores; (3) transport of both water and solute through the membrane pores is determined by interaction forces, friction forces, and chemical potential gradients of the water and solute; (4) the pores of the membrane are cylindrical and run the length of the membrane barrier layer; (5) a molecular layer of pure water is preferentially sorbed on the pore wall; and (6) a potential field controls the solute distribution of the membrane pore. An important modification of the SFPF model has been formulated which assumes a distribution of membrane pore sizes. Some inconsistencies in the SFPF model similar to those recognized for the original finely porous model (33,34) led to the formulation of a modified SFPF correcting these conceptual errors. The SFPF models require the solution of a velocity profile equation in the pore, force balance equation of solute in the pore, and correlations of interaction parameters with potential function. The pure water flow rate has been described by the Poiseuille equation.

The transport equations for the SFPF model, expressed in dimensionless form, have been solved using a variety of numerical techniques. Liquid chromatography (qv) was employed to determine F , measure of electrostatic repulsion force between the ionic solute and the membrane, or $B\sim$, the measure of short-range van der Waals forces for a solute (7,33,35). Trial-and-error was then used to find the membrane pore radius, R_p . Alternatively, if R_p was specified, then F or $B\sim$ was varied to produce agreement in the predicted and measured permeate concentrations. Measured pore radius values and one experimental data point for permeate concentration was also employed to eliminate the need for trial-and-error solution of the transport equations (36). Both solution techniques indicate that the SFPF model gives excellent predictions of solute separation for a wide range of inorganics and organics under varying operating conditions. However, for some dilute organics that cause substantial decreases in water flux, the models do not adequately predict the water flux.

Concentration Polarization. Concentration polarization (CP) is the term used to describe the accumulation of rejected solute at the surface of a membrane, causing the solute concentration at the membrane wall to be much greater than that of the bulk feed solution. The enhanced solute concentration on the membrane surface causes a reduction of the water flux owing to increased osmotic pressure (see eq. 5), precipitation of sparingly soluble salts such as CaSO_4 , reduced permeate quality, etc. A dramatic reduction in water flux for a CaSO_4 solution owing to CP is shown in Figure 3 (37). As soon as the mixing was stopped, CP increased so that the wall concentration of CaSO_4 exceeded saturation and the flux dropped steadily owing to precipitate formation.

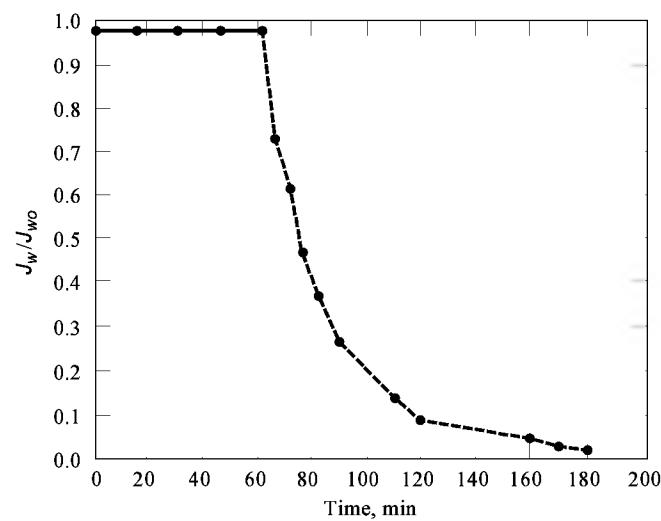


Fig. 3. Comparison of flux for an aromatic polyamide (FT30-BW) membrane for (—) mixed and (---) unmixed solutions of CaSO_4 , where the initial concentrations are 1.75 g/L and $\Delta P = 2.1 \times 10^6 \text{ N/m}^2$ (2.1 MPa (304.5 psi)). Mixing was carried out for 60 min (see eq. 4).

Reviews of concentration polarization have been reported (14,38,39). Because solute wall concentration may not be experimentally measurable, models relating solute and solvent fluxes to hydrodynamic parameters are needed for system design. The Navier-Stokes diffusion-convection equation has been numerically solved to calculate wall concentration, and thus the water flux and permeate quality (40).

A simplified model using a stagnant boundary layer assumption and the one-dimension diffusion-convection equation has been used to calculate wall concentration in an RO module. The integrated form of this equation, the widely applied film theory (41), is given in equation 8.

$$\frac{C_{\text{wall}}}{C_B} = \frac{C_P}{C_P} \exp\left(\frac{V_w}{k_s}\right) \tag{8}$$

Correlations for the mass-transfer coefficient, k_s , as the Sherwood number for various membrane geometries have been reviewed (39).

Using this simplified model, CP simulations can be performed easily as a function of solution and such operating variables as pressure, temperature, and flow rate, using software packages such as Mathcad. Solution of the CP equation (eq. 8) along with the solution-diffusion transport equations (eqs. 5 and 6) allow the prediction of CP, rejection, and permeate flux as a function of the Reynolds number, Re . To facilitate these calculations, the following data and correlations can be used: (1) for mass-transfer correlation, the Sherwood number, Sh , is defined as $Sh = 0.04 Re^{0.75} Sc^{0.33}$, where Sc is the Schmidt number; (2) osmotic pressure follows van't Hoff's equation, ie, $\pi = iC R_g T$, where i is the number of ions; (3) $A = 4.05 \times 10^{-7} \text{ cm}^3 (\text{s} \cdot \text{kPa}) / (4.1 \times 10^{-5} \text{ cm}^3 (\text{s} \cdot \text{atm}))$ and $B = 1.3 \times 10^{-4} \text{ cm}^3/\text{s}$; (4) $\mu = 1 \text{ mPa} \cdot \text{s}$ (cP), NaCl diffusivity in water $= 1.6 \times 10^{-5} \text{ cm}^2/\text{s}$, and solution density $= 1 \text{ g/cm}^3$. Figure 4 shows typical results of this type of simulation of salt water permeation through an RO membrane. Increasing the Reynolds number in Figure 4a decreases the effect of concentration polarization. The effect of feed flow rate on NaCl rejection is shown in Figure 4b. Because the intrinsic rejection, $R = 1 - C_P/C_{\text{wall}}$, is defined in terms of the wall concentration, theoretically R should be independent of the Reynolds number and in Figure 4b the intrinsic rejection is essentially so. On the other hand, as C_{wall} increases, C_P should increase and thus the observed rejection is expected to be a strong function of feed flow rate. The maximum flow rate that can be used to minimize CP is controlled by the membrane module pressure drop. The effects of spacers, complex flow patterns, types of fluids, etc on CP behavior have been reported (40).

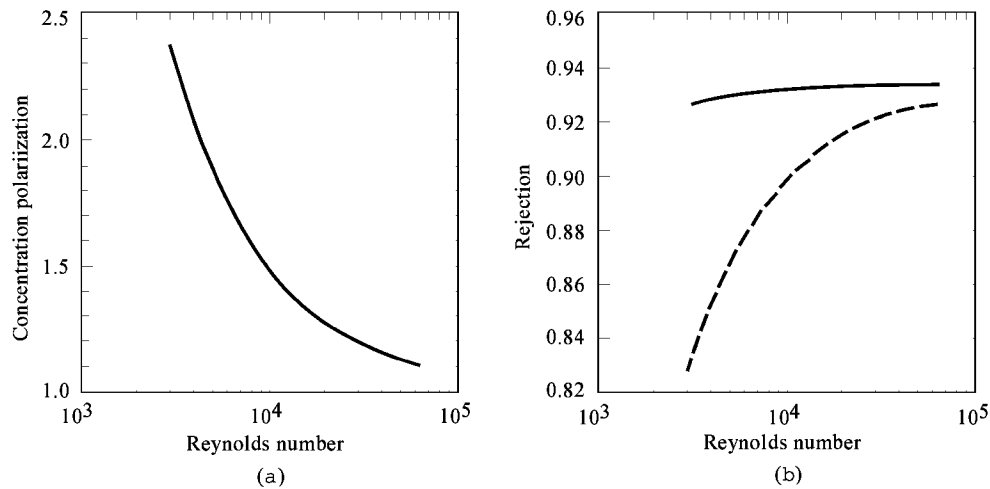


Fig. 4. Mathcad simulations ($C_F = 5000 \text{ mg/L}$) as a function of Reynolds number for a NaCl solution: (a) concentration polarization (CP), and (b) (---) observed and (—) intrinsic rejections using a simple film theory model.

Study of RO Variables and Typical Experimental Setup

Factors affecting RO membrane separations and water flux include feed variables such as solute concentration, temperature, pH, and pretreatment requirements; membrane variables such as polymer type, module geometry, and module arrangement; and process variables such as feed flow rate, operating time and pressure, and water recovery.

For simple, noninteracting inorganic salts, the trends in water flux and rejection behavior can be easily predicted (10). Most of the RO transport models (eq. 5) indicate that water flux, J_w , should increase linearly with the net pressure. Temperature affects variables such as viscosity and diffusivity. The result of a temperature increase is generally an increase in water (solvent) flux and a decrease in solute rejection. Water flux can also gradually decrease over operating time, measured in days or months of operation, as a result of fouling, membrane compaction, or other changes in membrane structure. Salt rejection (eq. 7) increases with the net pressure up to an asymptotic value. For ionizable organics, the rejection of ionized species is considerably higher than

the nonionized species.

The intrinsic rejection and maximum obtainable water flux of different membranes can be easily evaluated in a stirred batch system. A typical batch unit (42) is shown in Figure 5. A continuous system is needed for full-scale system design and to determine the effects of hydrodynamic variables and fouling in different module configurations. A typical laboratory pilot-scale continuous unit using computer control and on-line data acquisition is shown in Figure 6.

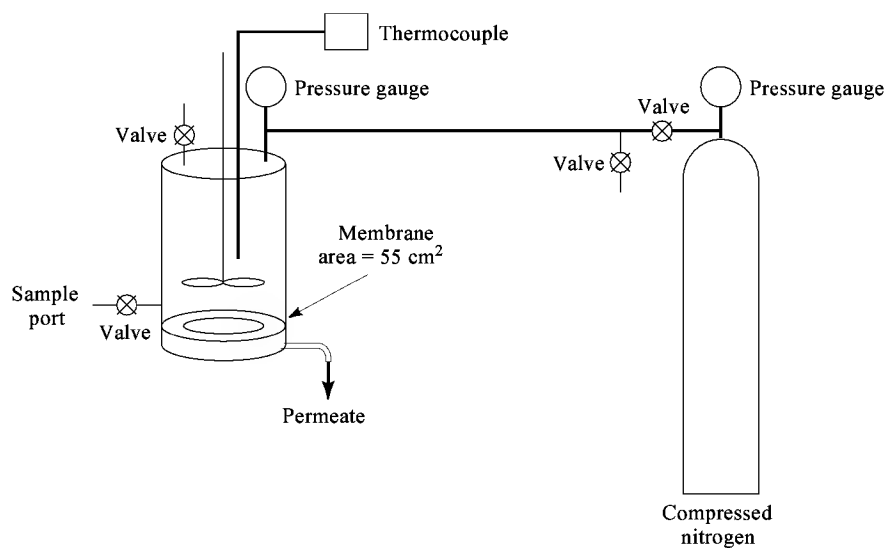


Fig. 5. Schematic for a stirred batch RO unit pressurized by compressed nitrogen.

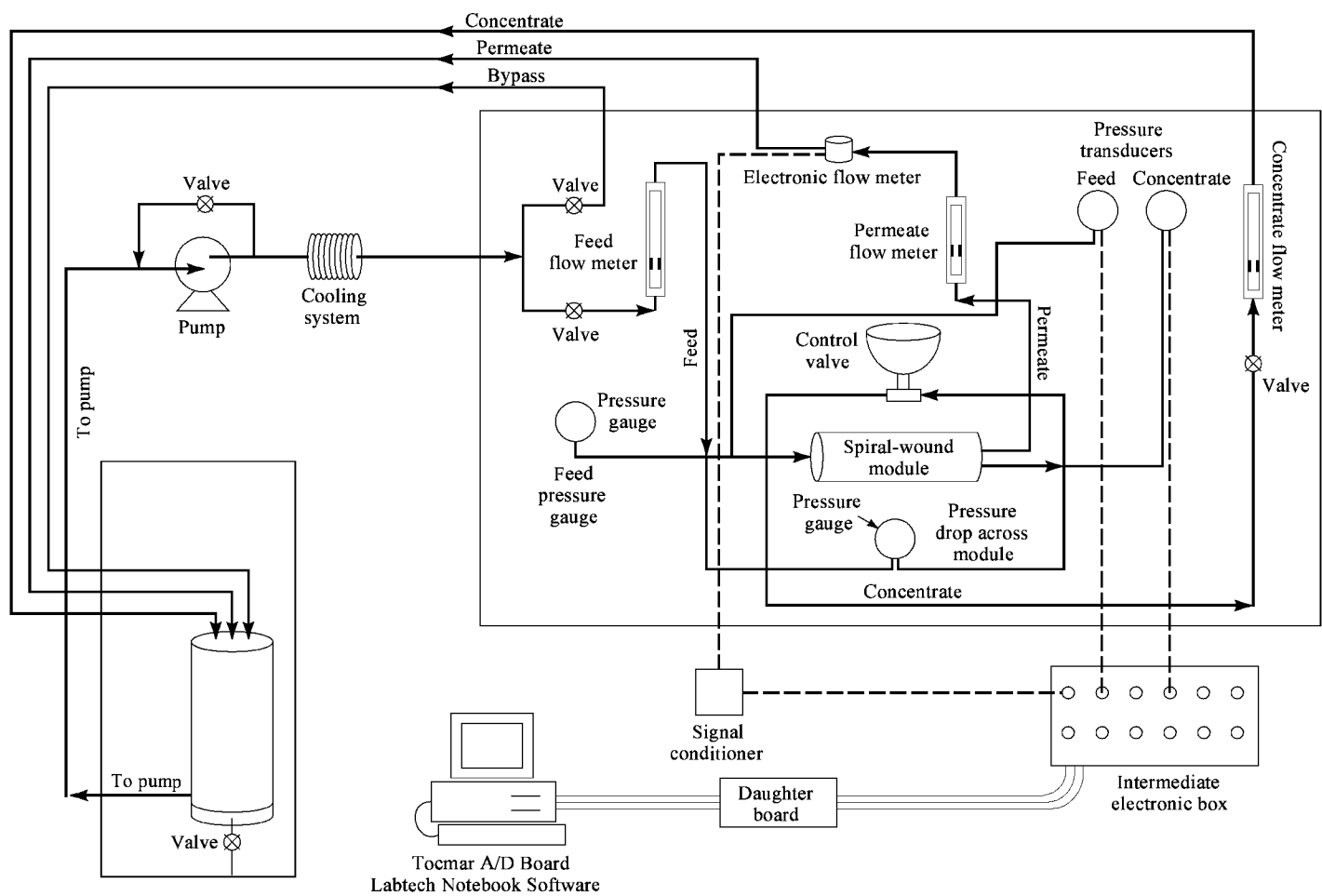


Fig. 6. Schematic for a computer-controlled, continuous RO unit pressurized by a pump. A/D = analog to digital.

Fouling. One of the primary concerns in reverse osmosis processes is the irreversible fouling of the membrane surface. This can lead to product flux decline as well as possible changes in rejection and selectivity. Fouling is a broad term that refers to the deposition or association of feed stream constituents on the membrane surface or within the membrane pores, consequently leading to a decrease in the membrane performance. The cause of fouling can generally be traced to constituents in the membrane feed stream. Several general categories of foulants found in process streams can be identified in relation to reverse osmosis processes: biological fouling agents, colloids (qv), scaling agents, organics, metal oxides, and suspended solids in general. Depending on the particular application, one or several of these categories may become a factor in the RO process. For example, in the desalination of seawater, problems with colloids, suspended solids, biological fouling, and scaling may occur. Wastewater reclamation, on the other hand, suffers primarily from fouling by dissolved organics and biological slime. Fouling must be taken into account when considering the use of RO on an industrial level.

The manner in which the foulants decrease the membrane flux varies, depending on the particular foulant. Gel layers can be formed at the surface from a variety of constituents that do not enter the membrane but rather accumulate at the surface. In the presence of inorganic salts at high water recovery, solubility limits can be exceeded, causing the formation of a scale on the membrane surface. Various microorganisms can attach and grow on the membrane, thus creating a biological slime. Other components that can form gel layers are humic acids, proteinaceous material, and other large molecules.

In the presence of fine colloids or suspended solids, the membrane pores may also become plugged.

The success of a reverse osmosis process hinges directly on the pretreatment of the feed stream. If typical process streams, without pretreatment to remove partially some of the constituents listed, were contacted with membranes, membrane life and performance would be unacceptable. There is no single pretreatment for all types of foulants. Pretreatment methods range from pH control, adsorption (qv), to filtration (qv), depending on the chemistry of the particular foulant. Some of the pretreatment methods for each type of foulant are as follow (43–45):

Fouling material	Pretreatment method
biological fouling	chlorination, ozonation, or uv light
colloids	coagulation filtration
scaling	acid treatment, antiscaling agents (SHMP), chelating agents (EDTA), or sand filtration
organics	coagulation filtration, activated carbon, or oxidation
metal oxides	proper materials or acid treatment
proteins	microfiltration or ultrafiltration
suspended solids	cartridge filtration or screen filters

Depending on the number, nature, and concentration of the potential foulants in the feed stream, a logical pretreatment plan must be developed to ensure the success of the RO plant.

Biologicals. Biological components form slimes on a membrane surface that can cause a reduction in the membrane flux as well as membrane degradation (43,45–47). Once microbes attach to the membrane surface, they can multiply. Steps such as chlorination (0.5 mg L) are generally taken to kill biological constituents before the feed enters the membrane module. Chlorine, however, causes damage to most RO membranes. Typical cellulose acetate membranes can withstand chlorine concentrations of 0.3–1.0 mg L; polyamide membranes can withstand less than 0.05 mg L. Therefore, any pretreatment that uses chlorine must have an additional step for chlorine removal, the most common of which is the use of sodium bisulfite. Too much residual chlorine after the removal step can lead to membrane damage, which in turn usually leads to an increase in flux accompanied by a dramatic decrease in rejection. If, however, all of the chlorine is removed while a few bacteria remain, the bacteria continue to grow. For this reason, it is desirable to have a chlorine-resistant membrane and prevent biological growth by allowing a low concentration of chlorine to be present in the RO module.

There has been considerable research on chlorine-resistant RO membranes (48–52). A poly(*trans*-2,5 dimethyl)piperazinthiofurazanamide used in the presence of low (3 mg L) concentrations of chlorine resulted in a membrane life of three years (48). A copolyamide hollow-fiber membrane for use in desalination has been developed that is resistant to 0.5 mg L chlorine (49). Millipore Corporation has also developed a sulfonated polysulfone member that has desirable chlorine-resistance properties.

Inorganics. Inorganic salts can precipitate on the membrane surface if the solubility limit of the salt is exceeded. This may occur upon increased water recovery. Depending on the types of salts present in the feed stream, different steps can be taken to prevent scaling. For waters containing bicarbonates and calcium, calcium carbonate, CaCO₃, presents a potential scaling problem. Treatment of the water with acid converts bicarbonate to carbon dioxide, thus eliminating the scaling problem (53). Calcium and magnesium can be removed by precipitation with lime and subsequent removal. Chelating agents (qv) such as ethylenediaminetetraacetic acid (EDTA) and polyphosphates are also frequently used to prevent precipitation.

Metal oxides found in RO feed streams typically originate from corroded pipes found in the RO process. These metal oxides can deposit on the membrane surface and decrease the membrane flux. This type of fouling can be prevented by using the proper materials of construction in the piping system to prevent corrosion.

Solids and Colloids. Suspended solids can accumulate at the membrane surface, creating an additional resistance to flow through the membrane as well as a possible feed channel, such as that for a spiral-wound module; plugging; and subsequently a decrease in flux. Prevention of this type of fouling lies in the removal of the suspended solids, which can be accomplished using filters and screens prior to arrival at the RO unit.

Colloidal materials present in surface waters can also plug RO membranes, causing a decrease in permeate flux. Colloidal plugging can be avoided by using one of several possible pretreatment steps. Ultrafiltration (qv) (UF) or microfiltration (MF), depending on the size of the colloid, can be used to filter out the colloidal material. Alternatively, a coagulant such as alum can be added to the water to form aggregates of the colloid, which can then be filtered in a similar manner as suspended solids.

Proteinaceous material is often found in process streams in the textile and food industries where RO is used to clean wastewaters (see FOOD PROCESSING; TEXTILES). The effect of proteinaceous material on membrane performance has been investigated extensively for ultrafiltration and microfiltration. Very little work has, however, been done in the RO area. Because the use of RO is growing in the area of wastewater treatment, there is a significant interest in determining the cause of proteinaceous fouling and methods of prevention. The size of protein molecules prevents them from entering the membrane matrix. Instead, these accumulate at the surface, forming a gel layer and creating an additional resistance to flow through the membrane. Depending on the isoelectric point(s) of the protein(s) in question, the pH of the feed stream, and the charge of the membrane, electrostatic interactions may cause association of proteins with the membrane. Another important factor that has been found to affect fouling in MF and UF is the presence of denatured protein (44), which may attach to the membrane surface to satisfy thermodynamic requirements. In MF and UF, surface modifications (qv) to impart hydrophilicity on the surface have been used to prevent fouling by proteinaceous material. For RO applications, MF and UF can be used as a pretreatment step to eliminate the protein at the RO module, but there may still be a fouling problem in the MF or UF modules.

Multiple Agents. In most applications, there is more than one foulant present in a given process stream. The best option in the operation of a reverse osmosis plant would be to eliminate fouling through module design and feed pretreatment. However, economic considerations almost always prevent the complete elimination of fouling. Typically, an economically feasible pretreatment process is designed, and when membrane performance degrades to a predetermined level, the process goes through cleaning cycles.

Membrane Cleaning. Methods for cleaning membranes have been developed (54–56). Cleaning methods for RO membranes (54) can be divided into two classes. Physical cleaning methods include the use of sponge balls, back-pressure flushing, and vibration. Chemical cleaning methods may employ EDTA or other chelating agents, surfactants (qv), acids, or sodium dithionate for iron-containing scales. Physical cleaning can only be performed on flat-plate (scrubbing) or tubular membranes (sponge balls). For the hollow-fiber and spiral-wound modules, which offer the advantage of high surface area, physical cleaning is not an option. Chemical treatments must be used to remove the fouling layer in these types of modules. For example, calcium scales can be removed by the use of citric acid. Once again, knowledge of the feed stream and hence the foulants is of the utmost importance when cleaning membranes. Extensive flux declines may also occur in systems containing dilute low molecular weight organics owing to interactions with membrane polymers.

Organic–Membrane Interactions

Understanding the feed characteristics of membrane polymers and organics solute interactions (57), fundamental transport mechanisms and simplified predictive models particularly for systems involving hazardous organics, and flux drop phenomena is essential for optimal RO design. Extensive experimental studies using various substituted, nonionized phenolic compounds have shown (42) that these materials can cause substantial membrane flux drop even in dilute solutions. These organics showed significant adsorption onto a cross-linked aromatic polyamide membrane during the RO operation.

Interactions. Solute–membrane interactions have been both measured and calculated. Liquid chromatography employing the membrane polymer as the column packing has been used to measure directly inorganic and organic solute–membrane interfacial parameters, such as equilibrium distribution coefficients, Gibbs free energy, and surface excess for a large number of different polymers (7,58,59). Additionally, the SFPF model has been used to calculate solute–membrane interaction forces, coulombic or van der Waals (7,33,36), and interactions of organic solutes with polybenzimidazole membranes have been characterized using quantum chemistry calculations, liquid chromatography, and ir spectroscopy (60).

The effects of small halocarbons, such as chloroform, bromoform, and carbon tetrachloride, on several RO membranes have been considered (61).

The organics were found to cause a reduction in the membranes' void spaces, representing a decrease in water content of the membrane, possibly because of substitution of the water in the membrane by the organic. Analysis of partition coefficients indicated that the organics were significantly sorbed into the membrane. Several other studies have also shown that even dilute solutions of some organic solutes, such as chloro- and nitrophenols, and polynuclear aromatic hydrocarbons interact with RO membranes, substantially changing water flux characteristics (9,62–64). Adsorption (qv) of the solute onto the membranes was thought to be related to the lower water flux observed.

Organic Rejections and Flux Drop Behavior. Many studies have been performed on the separation of organics and organic pollutants by RO membranes, and these studies have identified some of the unique aspects associated with organic separation. Separation and flux data of cellulose acetate membranes have been compiled for a large number of organic compounds, including many organic pollutants (7,65). Organic separation can vary from 0 to 100% , depending on the characteristics of the organic material, eg, polarity, size, and charge, and the RO operating conditions, such as feed pH and operating pressure.

Several investigations have compared organic rejections of cellulose acetate to that of other membranes. Aromatic polyamide and composite membranes usually have organic rejections greater than those of cellulose acetate membranes. Several organic rejections of the Toray composite membrane PEC-1000 (polyfuran) have been listed (66). Most rejections were high, including 97% for acetone and 99% for phenol. Separation results for several polar organic solutes, eg, alcohols, phenols, carboxylic acids, amines, and ketones, and various phenolic derivatives have been reported (67,68) for a composite membrane. The main factors affecting rejection included molecular weight, molecular branching, polarity, and degree of dissociation for ionizable compounds. Both cellulose acetate and thin-film composite membrane separations of a wide variety of organic pollutants have been compared (69), the latter involving FT30 (FilmTec), a cross-linked aromatic polyamide. The composite membrane rejections, greater than 90% for most of the organics studied, and water fluxes were substantially better than those of the cellulose acetate membrane. However, adsorption of some of the organics on the membranes was noted. Polyamide membrane rejections of chlorinated hydrocarbons (>95%) have been observed to be much greater than those of cellulose acetate membranes (70). Separation results for four composite and two asymmetric membranes for a variety of single and multicomponent organic solutions, including many organic pollutants, have been reported (71). Rejections varied from only 25% up to >99%, depending on the solute, but generally the composite membrane rejections were higher.

The rejection and flux characteristics of aromatic polyamide, FT30-BW, membranes for various chlorinated organics have been investigated (42). For substituted phenols, the rejections of ionized (pH > p*K_a*) species were always considerably higher than the nonionized species. Rejections of several environmentally significant organics are shown in Table 1. In general, rejection increased with increasing substituent group addition for the chloro- and nitrophenols. Both specific and nonspecific interactions are possible. On the other hand, water flux drop was highly significant for certain substituted, nonionized (pH < p*K_a*) organics. Figure 7 shows the normalized flux results for several organics. With the exception of PCP, the water flux drop caused by the dilute (insignificant osmotic pressure) chlorophenol solutions increased in the order of increasing numbers of chlorines on the phenol. The hydrogen bonding of the phenol to groups on the membrane polymer, such as the carbonyl group of polyamide FT30 membrane, could be expected to cause the observed flux drop by reducing the membrane hydrophilicity. Figure 8 supports this, showing a significant decrease in water flux as solute acidity (decreasing p*K_a*) increases. For these compounds, flux declines were insignificant when the RO system operated at pH values greater than the corresponding p*K_a*s, ie, the solutes were ionized.

Table 1. Rejections by Aromatic Polyamide Composite Membrane ^{a, b}

Compound	Rejection, %
standard NaCl	~98
nonionized phenolics	
nitrophenol	~55
phenol	~60
chlorophenol	~73
dinitrophenol	~80
trichlorophenol	~90
pentachlorophenol	~99
ionized phenolics ^c	94–99
other organics	
chloroaniline	>90
benzene	~93
dichlorobenzene	~96
ethyl benzene	>97
naphthalene	~98
trichlorobenzene	>99
anthracene	>99
phenanthrene	>99

^a RO membrane FT30-BW. Compounds are in dilute aqueous solution.

^b Pressure 1.3–2.0 MPa (13–20 atm).

^c For example, rejection of 2,4,6-trichlorophenol at pH 10 is 99.4% .

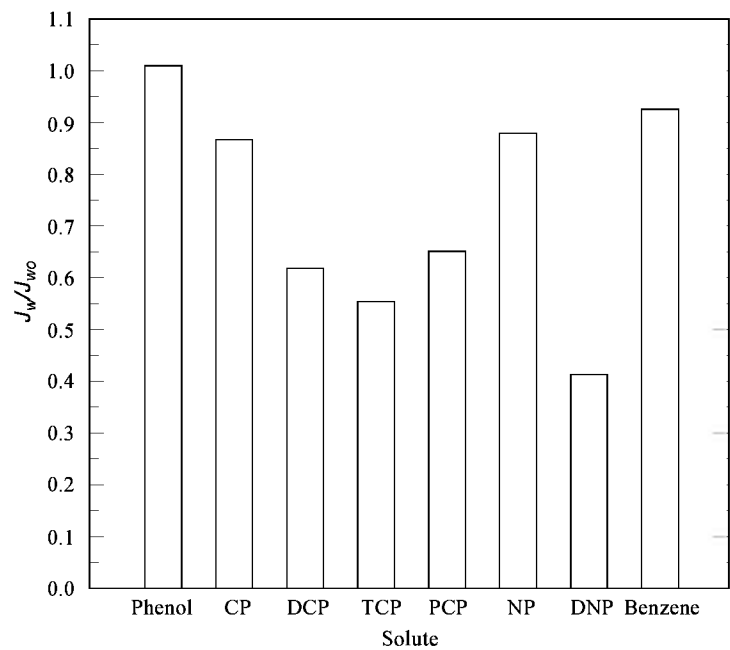


Fig. 7. Normalized water flux for hazardous organic compounds using FT30-BW membranes at $T = 24^{\circ}\text{C}$, $\Delta P = 1.38\text{ MPa}$ (200 psi), pH of the feed = 3.0–4.0, and $r_f = 52\text{--}62\%$. C_F for phenol, 2,4,6-trichlorophenol (TCP), and 2,4-dinitrophenol (DNP) was 0.24 mM; for 2-chlorophenol (CP), 0.23 mM; for 2,4-dichlorophenol (DCP), 0.22 mM; for 2-nitrophenol (NP), 0.21 mM; for benzene, 0.19 mM; and for pentachlorophenol (PCP), 0.02 mM.

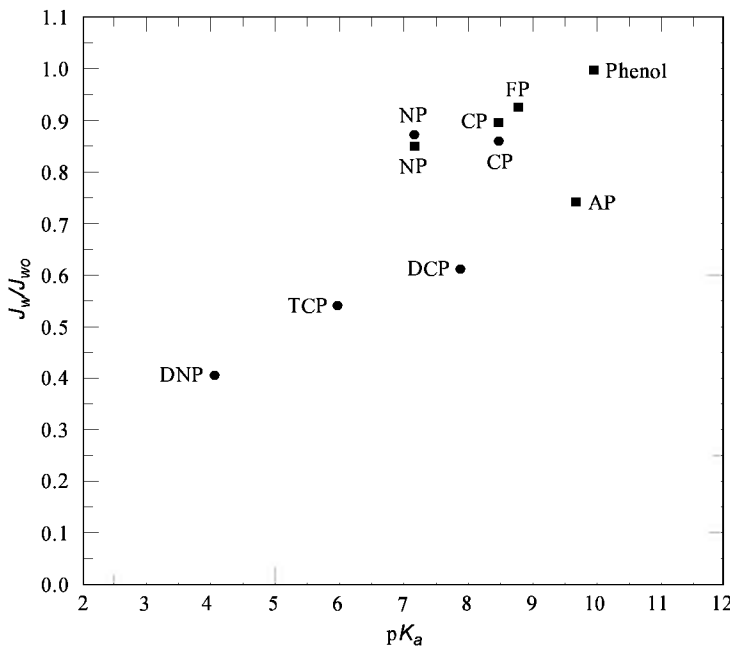


Fig. 8. Water flux as a function of pK_a for various phenols using FT30-BW membranes at $T = 24^{\circ}\text{C}$, $\Delta P = 1.38\text{ MPa}$ (200 psi), pH of feed = 3.0–4.6, and $r_f = 34\text{--}62\%$, where (■) represents $C_F = 0.5\text{ mM}$; (●), $\sim 0.2\text{ mM}$. AP = 2-aminophenol; FP = 2-fluorophenol; for other abbreviations, see Fig. 7.

A nonspecific interacting compound, benzene (see Fig. 7), caused a smaller ($<8\%$) water flux drop even though the benzene was more highly adsorbed ($2.9 \times 10^{-6}\text{ mol cm}^{-2}$) as compared to the specific interacting compound, 2,4,6-trichlorophenol (TCP), which was adsorbed to a lesser ($9.2 \times 10^{-7}\text{ mol cm}^{-2}$) extent but caused substantially more (almost 50%) flux decline. Flux and separations characteristics can be modeled upon modifying the SD equations by incorporating an organic sorption term (72) or by solving the general diffusion–convection equations using organic adsorption terms (42). A modified solution–diffusion (MSD) model can easily be obtained by assuming that the sum of organic and water concentrations is constant, ie, total concentration in membrane, $C_{tm} = C_m + C_{wm}$, and representing C_m/C_{tm} by a Langmuir adsorption term. The modified SD equations for water and solute flux are as in equations 9 and 10:

$$J_w = \left[\frac{1}{1 + b_0 C_F} \right] A^* (\Delta P - \Delta \pi) \tag{9}$$

$$J_s = B^* \left[\frac{b_0 C_F}{1 + b_0 C_F} - \frac{b_0 C_P}{1 + b_0 C_P} \right] \tag{10}$$

A simulation plot of normalized flux vs feed concentration, C_F is shown in Figure 9 for a known A^* , a modified water permeability parameter, and various values of b_0 , a measure of organic–membrane interaction parameter. Flux decreases significantly with increasing b_0 . From experimental results using FT-30 membranes, the following values of b_0 were obtained for nonionized organics: 280 M^{-1} for 2-chlorophenol, 1380 M^{-1} for 2,4,6-trichlorophenol, and 4400 M^{-1} for 2,4-dinitrophenol. Thus RO membranes typically used for treatment of aqueous feeds are not as effective in the treatment of organic feeds. Development of novel RO membranes for the separation of organics would be of interest.

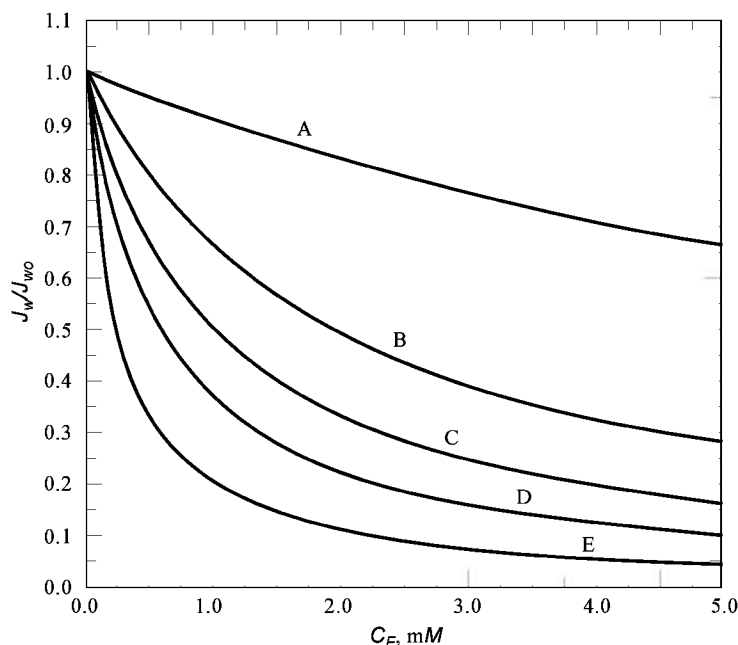


Fig. 9. Parametric simulation using the MSD model of water flux vs feed concentration as a function of organic sorption coefficient, b_0 , where for A, $b_0 = 100 M^{-1}$; for B, $500 M^{-1}$; for C, $1000 M^{-1}$; for D, $1695 M^{-1}$; and for E, $4000 M^{-1}$.

Applications

Developments and advances in both membrane materials and reverse osmosis modules have increased the range of applications to which RO can be applied. Whereas the RO industry has developed around water desalination (9,53,73,74), RO has become a significant cornerstone in other industries. These include wastewater cleanup for electroplating (75–78), radioactive processing (79–82), landfill leachate (76,83), and municipal wastewater (84–87); ultrapure water production for electronics-grade (88,89), laboratory-grade (90), and pharmaceutical-grade (91) materials; and food processing (qv) (9).

Wastewater Applications. Significant advances have been made to reduce environmental pollution problems caused by the generation of hazardous wastes. Waste management approaches include abatement, minimization, recycle and reuse, end-of-pipe treatment, and environmentally safe disposal of effluents and residuals (see *WASTES, INDUSTRIAL*). Even using the implementation of waste reduction procedures, significant amount of wastes often require further treatment. These treatment technologies may include biological treatment, adsorption, stripping, chemical and photolytic oxidation, membrane separation (RO, NF, pervaporation, catalytic membranes), and supercritical water oxidation. For separation, material recovery, and destruction of hazardous organics, a combination of membrane processes with other technologies can provide the optimum treatment system. Many wastewater treatments use reverse osmosis.

Several advantages of the RO process make it particularly attractive for dilute aqueous wastewater treatment: (1) RO systems are simple to design and operate, have low maintenance requirements, and are modular in nature, making expansion of the systems easy; (2) both inorganic and organic pollutants can be removed simultaneously by RO membrane processes; (3) RO systems allow recovery/recycle of waste process streams but do not affect the material being recovered; (4) RO membrane systems often require less energy and offer lower capital and operating costs than many conventional treatment systems; and (5) RO processes can considerably reduce the volume of waste streams so that the resulting concentrated system can be treated more efficiently and cost-effectively by other processes, such as incineration or other destruction processes (75,92,93) (see *WASTE TREATMENT, HAZARDOUS WASTES*). In addition, RO systems can replace or be used in conjunction with other treatment processes, including oxidation, adsorption, stripping, or biological treatment, to produce a high quality product water that can be reused or discharged.

Applications reported for RO processes include the treatment of organic-containing wastewater, wastewater from electroplating and metal finishing, pulp and paper, mining and petrochemical, textile and food processing industries, radioactive wastewater, municipal wastewater, and contaminated groundwater (75,76,83,94–99). One of the newer applications involves the treatment of wastewater generated by the semiconductor industry, which generates wastewater containing large amounts of hydrofluoric acid (HF) as well as ionic impurities that render the HF solution useless. RO membranes that allow selective permeation of HF while rejecting the ionic impurities to produce a pure HF solution that can be reused have been reported (100).

Electroplating and Metal-Finishing Process Wastewaters. In most cases, process wastewaters from the electroplating (qv) and metal-finishing industries (see *METAL SURFACE TREATMENT*) must be treated to remove heavy metals before being discharged. Reverse osmosis is ideal for many of these operations because it allows both recovery of the heavy metals and reuse of the product water in the process. The RO process has been used in the treatment and recovery of wastewater containing nickel, acid copper, zinc, copper cyanide, chromium, aluminum, and gold (75,97,98). Results for an industrial printed circuit board production plant have been summarized (77), demonstrating the effectiveness of RO for plating salts recovery and virtual elimination of wastewater discharge.

Cellulose acetate and polyamide (FT30) membranes were used at three Japanese plating shops employing nickel, chromium, and gold plating lines (78). Up to 80% water recoveries and high metal and total dissolved solids (TDS) (>95%) rejections were possible. The product water was recycled. The RO processes were found to be cost-effective in treating the wastewaters, and the compact nature of the RO system made it highly desirable to the customers because of space limitations. The RO treatment of effluent from an electrolytic polishing process for aluminum products has also been discussed (101). The streams contained phosphoric acid and aluminum from rinse water. DDS HR-98 membranes allowed 96–98% acid recovery, up to an acid concentration of 20%, and produced permeate water suitable for reuse. The membranes appeared to be stable to the feed even at the low (0.9–1.0) pH values found at high recoveries.

Radioactive Processing Wastewaters. Because of high rejection of inorganic compounds, RO membranes have been studied for treatment of radioactive effluents. These membranes have been used to treat uranium conversion process effluent containing toxic, corrosive, and radioactive compounds (79). The FT30 membranes had rejections of uranium of 99.5% for water recoveries up to 70%. Results indicated that the treated effluent was within regulatory discharge standards. In addition, a hybrid process consisting of nanofiltration (NF), reverse osmosis, and precipitation has been used to treat uranium effluents (80). The process removed both soluble and suspended uranium species. Up to 95% uranium recovery was possible, and the treated effluent met environmental standards. The FT30 membranes gave uranium rejections of >99%. Cellulose acetate membranes have been shown to remove 99% or uranium from effluents containing uranium nitrate when the uranium was complexed with EDTA (81). RO can be used to treat contaminated effluent from a uranium metal plant (82), whereby 95% by volume of the treated effluent is within discharge limits.

Leachates. The treatment of landfill leachates serves two purposes. It facilitates the production of a water stream that can be discharged while concentrating the hazardous material and reducing its volume. Several studies have been conducted on the treatment of landfill leachates using RO processes. Reverse osmosis has been shown to be effective, ie, high rejection at high recovery, in the removal of total organic carbon (TOC), TDS, chemical oxygen demand (COD), ammonia, heavy metals, and alkalinity from different types of leachates (102,103). In some cases, flux decline has been

reported. However, proper pretreatment and membrane cleaning can alleviate this problem. In an early study, RO membranes were used to remove >91% of TOC from sanitary landfill leachate (102). In one case, RO treatment of landfill leachate resulted in removals of 94.5% alkalinity, 97% COD, 97% total solids, 92.1% volatile solids, and 96.6% ammonia (104). FT30 membranes were used to treat soil-wash leachates containing heavy metals and organic contaminants so that the treated water could be recycled back to the soil-washing step (64). TOC rejections as high as 80–85% and heavy metal (Pb, Zn, Ni, Cu) rejections of 94–98% were found. However, water flux decreases of up to 33% were noted. The effects of the addition of EDTA or surfactant and feed preozonation were also investigated. Feed preozonation was found to have substantially improved membrane water flux. Specific organic rejections included >98% for pentachlorophenol and 2,4-dinitrophenol, >97% for ethylbenzene, >81% for xylene, and >90% for chloroaniline. The treatment of landfill leachates containing organic acids has been reported (105). Good separations of volatile fatty acids were obtained and TDS removed sufficiently to allow discharge of the product water. A commercial leachate treatment site in Germany was operational for >10,000 h without membrane replacement (83). Tests on leachates from sites where the waste was segregated have been conducted (76). Waste containing a large amount of ash was found to be unsuitable for treatment by RO.

Drinking Water. The ability of RO membranes to remove both inorganic and organic compounds have made these systems attractive for the treatment of contaminated drinking water supplies (106). Reverse osmosis processes can simultaneously remove hardness, various inorganics such as bromate and bromide ions (107), color (108), many kinds of bacteria and viruses, and organic contaminants such as agricultural chemicals and trihalomethane precursors. RO treatment of drinking water sources has been reviewed (109), and successful removal of a wide variety of contaminants has been accomplished. Membrane processes are also finding applications for the removal of natural organic matter (NOM) and disinfection by-products (DBPs). An excellent overview of selected processes, eg, adsorption, coagulation, RO, and NF, for removing NOM and DBPs has been reported (110). Membrane processes show the potential to achieve the highest removal of NOM and DBP precursors, but for surface water applications, high pretreatment are required to minimize fouling.

Municipal Wastewater. The application of RO membranes to the treatment of municipal wastewater has also had some success (see WATER, MUNICIPAL WATER TREATMENT). Reverse osmosis can remove dissolved solids that cannot be removed by biological or other conventional municipal treatment processes. In addition, RO membranes can also lower organics, color, and nitrate levels. However, extensive pretreatment and periodic cleaning are usually needed to maintain acceptable membrane water fluxes. Early studies showed that high removals of TDS and moderate removals of organics could be achieved. Use of tubular membranes and a substantial pretreatment system of the process stream removes both inorganics and organics from municipal secondary effluent and produces water meeting drinking water standards (84).

The Water Factory 21 site in Orange County, California, is a large-scale municipal wastewater treatment site that has been studied in detail (85,86). The feed to the plant consisted of secondary effluent, and the process was composed of a variety of treatment systems, including several different types of RO membranes having a $1.9 \times 10^4 \text{ m}^3 \text{ d}$ capacity. The process reduced TDS and organics to levels that allowed the effluent to be injected into groundwater aquifers used for water supplies. Use of several RO membranes to treat secondary effluent containing various salts and dissolved organic materials has been reported (87). TDS rejections of up to 99% and TOC rejections as high as 90% were found possible, and fecal coliform group rejections were >99.9%. Losses in water flux over time were noted but these could be partially restored by periodic cleaning.

Desalination. Desalination of seawater and brackish water has been and, as of the mid-1990s, is the primary use of RO. Driven by a need for potable water in areas of the world where there is a shortage, this industry has developed. Desalination involves the reduction of the total dissolved solids (TDS) concentration to less than 200 mg/L. RO offers several advantages over other possible desalination processes such as distillation (qv), evaporation (qv), and electrodialysis. The primary advantage of RO over the traditionally used method of distillation is the energy savings that is afforded by the lack of a phase change in RO.

Reverse osmosis processes for desalination were first applied to brackish water, which has a lower TDS concentration than seawater. Brackish water has less than 10,000 mg/L TDS; seawater contains greater than 30,000 mg/L TDS. This difference in TDS translates into a substantial difference in osmotic pressure and thus the RO operating pressure required to achieve separation. The need to process feed streams containing larger amounts of dissolved solids led to the development of RO membranes capable of operating at pressures approaching 10.3 MPa (1500 psi). Desalination plants around the world process both brackish water and seawater (15).

Most desalination plants use similar pretreatment and post-treatment methods in the desalting process (Fig. 10). Feed water is passed through multimedia, sand, and cartridge filters to remove particulates and suspended solids. In addition, the feed water is typically treated with chlorine to kill any microorganisms in the water, followed by coagulation and filtration to remove the dead cells. For most commercially available membranes, dechlorination using sodium bisulfite before entering the membrane module is necessary to prevent membrane damage. Chelating agents and acid are injected into the feed water to prevent precipitation and scaling on the membrane surface. Post-treatment of the permeate usually consists of the addition of lime to increase the pH along with chlorine to prevent biological growth. In seawater desalination, the concentrated brine is usually returned to the source. For brackish and well water desalination, disposal of the brine is necessary.

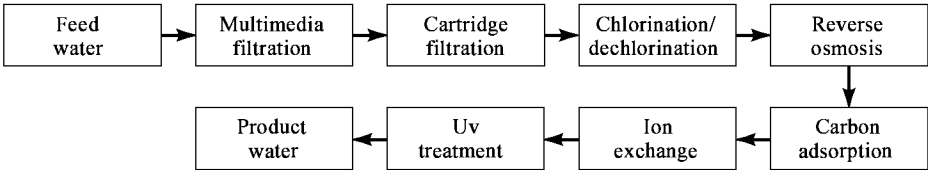


Fig. 10. Flow diagram for the setup of a desalination plant employing reverse osmosis, including pre- and post-treatment steps.

The largest desalination plant in the world is a brackish water plant in Yuma, Arizona having a capacity of 275,000 m³/d (73). Spiral-wound cellulose acetate membranes produce permeate water having 200 mg/L TDS from intake water containing 3100 mg/L TDS. Seawater desalination requires a much higher pressure to achieve a given separation than brackish water. The largest seawater desalination plant in the world operates in Jeddah, Saudi Arabia. It has a capacity of 56,800 m³/d, and the TDS content of the seawater is approximately 44,000 mg/L (74). Pretreatment modules for this plant include chlorine treatment, a dual media filter, and a cartridge filter.

Some desalination plants combine distillation with reverse osmosis to produce both power and water. Multistage flash (MSF) processes are used to produce both power and distilled water. The combination of RO and MSF and the advantages of such a combination have been reported (111). Distillation processes typically reduce the TDS concentration to levels well below the required specifications. Because the product water from the two processes is combined, the RO process can produce water at higher TDS concentrations and still meet the potable water specifications. In addition, the power produced from the MSF process can be used in the RO process, cutting energy costs.

Ultrapure Water. Reverse osmosis has been used to produce ultrapure water for a variety of applications. The electronics industry requires extremely pure water for the manufacture of semiconductors (qv) and other electronic components (88,89) (see ELECTRONIC MATERIALS). For example, sodium should be present in no greater concentration than 0.001 mg/L and there can be no more than one microorganism per mL. Whereas RO removes the low molecular weight organics that pass through an ultrafiltration membrane, by itself, RO cannot produce water of this purity. Rather, RO is merely one step in a hybrid of processes designed to achieve this level of purity. The development of new membrane modules for ultrapure water production processes has been reported (112). These membranes showed high rejections of TOC and inorganic ions at low concentrations.

RO is also used to produce ultrapure water for many laboratory uses (90) as well as in the medical and pharmaceutical industries (91). As for the electronics industry, purity is achieved using a combination of processes. A typical hybrid process for the production of ultrapure water is shown in Figure 11. The order in which the various steps take place may vary from case to case.

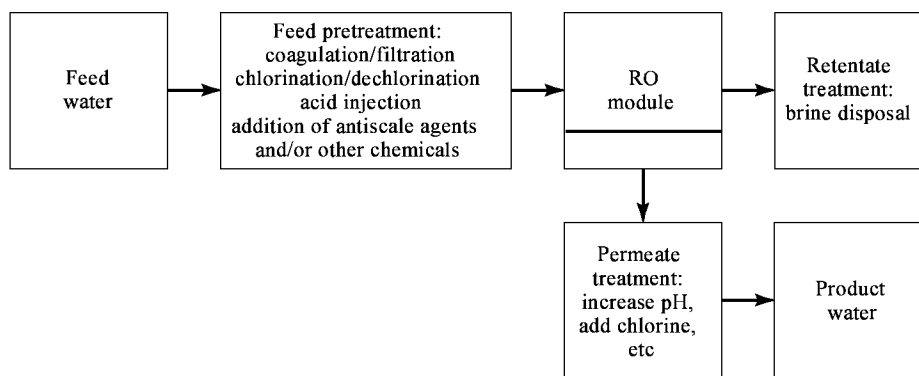


Fig. 11. Schematic of membrane-based hybrid process for ultrapure water production.

Food Processing. One of the first applications of RO was in the food processing (qv) industry. The primary advantage of RO over the traditionally used processes in the food industry is that RO operates at low temperatures which can prevent the denaturation of some materials used in foodstuffs. Because high temperatures are not required, energy costs are reduced as well. Moreover, RO is relatively simple in terms of the equipment design. These factors lead ultimately to a reduction in capital and operating costs, accompanied by an increase in product quality.

RO, primarily used in the dairy industry, is expanding into other areas of food processing. RO can be used for a variety of operations, ranging from wastewater treatment and material recovery to clarification and concentration. Material recovery is advantageous for two reasons. By recovering valuable products, eg, proteins, from waste streams, profits can be increased while costs for waste disposal decreased. An excellent review of the different applications of RO in food processing is available (9).

Nanofiltration

Nanofiltration, also called loose RO, membranes are a more recent development in the field of RO membrane separations. These membranes typically have much higher water fluxes at low pressures when compared to traditional RO membranes. Nanofiltration (NF) membranes are usually charged, ie, have carboxylic groups, sulfonic groups, etc, and as a result, ion repulsion (Donnan exclusion) is the primary factor in determining salt rejection. More highly charged ions such as SO_4^{2-} are rejected more fully than monovalent ions such as Cl^- by a negatively charged nanofiltration membrane. Another characteristic of charged membranes is the tendency to foul in the presence of components having opposite charge to that of the membrane (113).

Nanofiltration membranes usually have good rejections of organic compounds having molecular weights above 200–500 (114,115). NF provides the possibility of selective separation of certain organics from concentrated monovalent salt solutions such as NaCl . The most important nanofiltration membranes are composite membranes made by interfacial polymerization. Polyamides made from piperazine and aromatic acyl chlorides are examples of widely used nanofiltration membrane. Nanofiltration has been used in several commercial applications, among which are demineralization, organic removal, heavy-metal removal, and color removal (116).

Nanofiltration has been used to remove both organics and inorganics in various wastewaters. The use of NF membranes to remove color-causing compounds from effluent containing lignins and high salt concentrations in a wood pulping process has been reported (117). Color removals were >98% at water recoveries up to 95%, whereas the inorganics were poorly rejected, allowing the use of low operating pressures. NF can give high separations of color-causing compounds such as lignin sulfonates in paper pulping wastewaters (118). NF removes >95% of chlorinated organic compounds from alkaline pulp and paper bleaching effluents having high water fluxes (119). The use of NF membranes to remove hardness and organics in textile mill effluents (120), the recovery of low molecular weight dyes from high salt concentration effluent (121), and the use of NF membranes in the treatment of food processing wastewaters (115,118) have been reported. Specific uses included the desalting of whey and the reduction of high biological oxygen demand (BOD) and nitrate levels in potato processing waters (122).

NF processes are also finding wide applications in drinking water production by removing natural organics from water that cause trihalomethane formation during coagulation (116). NF membranes have been used to remove organic matter from highly colored groundwater (123) as well as from river water (124). Nanofiltration has been used in a hybrid process employing biological treatment and chemical oxidation to treat leachate from a dumpsite (125). Increases of 9–17% in the elimination rate as compared to a process without nanofiltration have been observed.

NF40 (Dow FilmTec) membranes were used to separate mixtures of cadmium and nickel (126). These membranes have also been examined with and without feed pretreatment by ozonation (127,128) for removal of various chlorophenols and chloroethanes. TOC rejections up to 90% were possible using ozonation pretreatment. The use of NF in a process for treating uranium wastewater where NF40 uranium rejections were 97–99.9% has been reported (80). The use of NF membranes to replace activated carbon filters for the removal of organics from offshore-produced water containing residual oils has also been discussed (129). The produced waters contained ~1000 mg/L soluble organics, mostly carboxylic acids, and high inorganic concentrations, ~15,000 mg/L Na^+ and ~25,000 mg/L Cl^- , as well as other dissolved ions. Organic rejections were suitable to meet discharge standards; inorganic rejections were low (<20%), allowing operation at low pressures.

Design Considerations

To ensure the successful design of a reverse osmosis process, several factors should be considered. These considerations encompass the feed solution, the membrane module, and the use of other processes in the pre- and post-treatment processes. A thorough knowledge of the feed stream and its components is essential to the prevention of membrane damage and product impurities. Once the feed stream is characterized and the process objective is defined, design can be initiated.

Feed characterization, particularly for nondesalination applications, should be the first and foremost objective in the design of a reverse osmosis plant. This involves the determination of the type and concentration of the main solutes and foulants in the stream, temperature, pH, osmotic pressure, etc. Once the feed has been characterized, a realistic process objective can be defined. In most cases, some level of pretreatment is needed to reduce the number and concentration of foulants present in the feed stream. Pretreatment necessitates the design of processes other than the RO module, thus the overall process design should use the minimum pretreatment necessary to meet the process objective. Once the pretreatment steps have been determined and the final feed stream defined, the RO module can be selected.

Design of the membrane module system involves selection of the membrane material; the module geometry, eg, spiral-wound or hollow-fiber; product flow rate and concentration; solvent recovery; operating pressure; and the minimum tolerable flux (9,11). The effects of these variables can be obtained from laboratory or pilot experiments using different membranes and modules. The membrane module as well as the solvent recovery can be chosen to minimize fouling. Spiral-wound modules are widely used because these offer both high surface area as well as a lower fouling potential.

Prediction of reverse osmosis performance is useful to the design of RO processes. Simulation of RO processes can be separated into two categories. The first is the prediction of membrane module performance. The second is the simulation of a network of RO processes, ie, flow sheet simulations, which can be used to determine the optimum placement of RO modules to obtain the overall process objective.

Owing to the variety of situations encountered in RO applications, there is no single analytical technique to predict membrane module performance. The module and the feed stream, along with the operating parameters, determine system performance. To predict module performance, a model that

governs transport through the membrane must be used. A wide range of models have been proposed, ranging from the solution–diffusion model to the irreversible thermodynamics model. Each model is based on several assumptions and is therefore limited in its range of applicability. Simulations that make use of these models have been developed for several systems. Data obtained from a pilot plant in the treatment of cooling tower water by tubular reverse osmosis have been used to simulate a full-scale plant (130). The finely porous model for transport through the membrane is employed. The solution–diffusion model has also been used to simulate the transport of a saline solution through the DuPont B10 permeator (131). A finite-element technique was used (40) to predict the concentration polarization in a parallel-plate membrane module, which involves the solution of convection–diffusion equation, given a mass-transfer model and boundary conditions. This model was used to determine the effect of the operating variables on the flux and rejection behavior of NaCl systems.

Given the first type of simulation, it is advantageous to be able to design a system of RO modules that can achieve the process objective at a minimal cost. A model has been integrated into a process simulation program to predict the stream matrix for a reverse osmosis process (132). In the area of waste minimization, the proper placement of RO modules is essential for achieving minimum waste at a minimum cost. Excellent details on how to create an optimal network of RO modules is available (96).

Economic Aspects

Economics plays a role in the design of any industrial process and RO is no exception (133–136). In 1990 sales of \$118 million were attributed to the RO industry (10). Costs of RO and evaporative processes involving desalination of different feed water sources have been compared (133). RO was found to be the lowest in cost compared to multistage flash and multieffect evaporator having thermal vapor compression. The costs associated with RO processes have been reviewed (134). Capital costs include equipment, land, installation, and utilities. Equipment costs include the components for pretreatment of the feed, storage of the product, pipes, pumps, and the RO module. Equipment costs in terms of capacity corresponding to 1 m³ d range from \$400 for brackish water plants to \$900 for seawater plants. Operating costs include the usual costs for an industrial process: energy, labor, and chemicals. However, for the RO process there are additional operating costs from membrane replacement, which range from \$0.05 m³ for brackish water plants to \$0.10 m³ for seawater plants. Energy consumption in seawater desalination reverse osmosis is usually the largest operating cost because of the high pressures required. Energy costs have been reported from \$0.10 m³ to \$0.40 m³.

Overall production costs range from \$0.50–1.85 m³, depending on the type, size, and location of the plant. Other factors contributing to costs are the extent of pretreatment necessary for acceptable performance, and the degree of solvent recovery. For applications other than desalination, economic calculations must also include the processing of retentate (the concentrate stream) and the benefits of material recovery and water reuse. If valuable materials can be recovered by an RO process, that added value decreases the total production costs. The integration of RO processes and other concentration methods, such as evaporation and pervaporation, can provide significant minimization of waste by simultaneous material recycling and water reuse.

The primary products of the RO industry are the membrane modules that determine both the geometry and specific membrane type. Therefore, the success of the RO industry is dependent on the production of inexpensive, long-lasting membrane modules. These modules must be mechanically stable at high pressures and provide good hydrodynamics to minimize fouling. Maximizing the membrane surface area per unit volume of the module is also a critical objective. The predominant modules used industrially are the spiral-wound module and the hollow-fiber module, both of which offer high membrane surface area. Costs and capacities of spiral-wound and hollow-fiber thin-film composite membrane modules are given in Table 2.

Table 2. Capacities and Costs of Desalination Processes^a

Desalination feed	Capacity, m ³ d	Cost, \$ ^b
	<i>Hollow-fiber module</i>	
seawater	18.9	322
brackish water	60.6	63.7
	<i>Spiral-wound module</i>	
seawater	15.1	101.3
brackish water	28.4	42.3

^a The membrane material is a thin-film composite in a 20-cm module.

^b For each m³ d.

RO membrane modules are available from many manufacturers including, for hollow-fiber modules, DuPont and Dow FilmTec Corporation, and for spiral-wound modules, UOP Inc., Millipore Corporation, Nitto-Denko America, Inc., Toray Industries Inc., Dow FilmTec Corporation, and DuPont.

NOMENCLATURE

Symbol	Definition	Units
A	water permeability coefficient	$L^3 (L^2 \cdot t P)$
A^*	modified water permeability constant	$L^3 (L^2 \cdot t P)$
Γ	measure of electrostatic force between solute and pore wall	L
A_m	membrane surface area	L^2
B	solute permeability coefficient	t^{-1}
$B\sim$	measure of solute–membrane pore wall interaction force	L^3
B^*	modified solute permeability constant	$\text{mol} (L^2 \cdot t)$
b_0	sorption coefficient	$L^3 \text{ mol}$
C_B	bulk concentrate solute concentration	$\text{mol } L^3 \text{ or } M L^3$
C_F	feed solute concentration	$\text{mol } L^3 \text{ or } M L^3$
C_m	membrane solute concentration	$\text{mol } L^3$
C_P	permeate solute concentration	$\text{mol } L^3 \text{ or } M L^3$
C_{tm}	total concentration of water and solute in the membrane	$\text{mol } L^3$
C_{wall}	feed solute concentration at the membrane wall	$\text{mol } L^3 \text{ or } M L^3$
C_{wm}	membrane water concentration	$\text{mol } L^3$
D_{tm}	membrane solute diffusivity	$L^2 t$
D_{wm}	membrane water diffusivity	$L^2 t$
E	electrochemical potential	mV or V
F_F	feed flow rate	$L^3 t$
F_P	permeate flow rate	$L^3 t$

J_s	solute flux	$\text{mol } (L^2 \cdot t) \text{ or } M \text{ } (L^2 \cdot t)$
J_w	water flux	$L^3 \text{ } (L^2 \cdot t)$
J_{wo}	pure water flux	$L^3 \text{ } (L^2 \cdot t)$
K_m	solute partition coefficient for homogeneous membrane	dimensionless
k_s	mass-transfer coefficient	$L \text{ } t$
L	length	cm or m
P	pressure	kPa or Pa
ΔP	transmembrane pressure difference	$M \text{ } (L \cdot t^2)$
r	permeate water recovery	dimensionless
R	solute rejection	dimensionless
Re	Reynolds number	dimensionless
R_g	gas constant	$M L^2 \text{ } (t^2 \cdot T) \text{ mol}$
R_p	membrane pore radius	L
Sc	Schmidt number	dimensionless
Sh	Sherwood number	dimensionless
t	time	s
T	temperature	K
V_F	feed volume	L^3
V_P	permeate volume	L^3
V_w	water permeation velocity	$L \text{ } t$
$\overline{V_w}$	partial molar volume of water	$L^3 \text{ mol}$
δ	membrane thickness	L
$\Delta \pi$	transmembrane osmotic pressure difference	$M \text{ } (L \cdot t^2)$
μ_s	solute chemical potential	$E \text{ mol}$
μ_w	water chemical potential	$E \text{ mol}$
σ	Staverman reflection coefficient	dimensionless

BIBLIOGRAPHY

"Osmosis and Osmotic Pressure" in *ECT* 1st ed., Vol. 9, pp. 643–660, by E. H. Immergut and K. G. Stern, Institute of Polymer Research, Polytechnic Institute of Brooklyn; "Osmosis, Osmotic Pressure, and Reverse Osmosis" in *ECT* 2nd ed., Vol. 14, pp. 345–356, by B. Keilin, Amicon Corp.; "Reverse Osmosis" in *ECT* 3rd ed., Vol. 20, pp. 230–248, by J. S. Johnson, Jr., Oak Ridge National Laboratory.

1. C. Reid and E. Breton, *J. Appl. Polym. Sci.* **1**, 133 (1959).
2. P. Ferguson, *Desalination*, **32**, 5 (1980).
3. H. Lonsdale, *J. Membrane Sci.* **10**, 81 (1982).
4. L. Applegate, *Chem. Eng.* **91**, 64 (June 11, 1984).
5. S. Loeb and S. Sourirajan, *Adv. Chem. Series*, **38**, 117 (1962).
6. S. Loeb, in A. Turbak, ed., *Synthetic Membranes*, American Chemical Society, Washington, D.C. (1981).
7. S. Sourirajan and T. Matsuura, *Reverse Osmosis/Ultrafiltration Principles*, National Research Council of Canada, Ottawa, Canada, 1985.
8. R. J. Petersen, *J. Membrane Sci.* **83**, 81–150 (1993).
9. D. Bhattacharyya and co-workers, in W. Ho and K. Sirkar, eds., *Membrane Handbook*, Van Nostrand Reinhold Co., Inc., New York, 1992, pp. 263–390.
10. R. Riley, in *Membrane Separation Systems*, Vol. 2, U.S. DOE Report, DOE ER 30133-H1, U.S. Dept. of Energy, Washington, D.C., 1990.
11. B. Parekh, ed., *Reverse Osmosis Technology*, Marcel Dekker, Inc., New York, 1988.
12. G. Belfort, ed., *Synthetic Membrane Processes: Fundamentals and Water Applications*, Academic Press, Inc., New York, 1984.
13. D. Lloyd and T. Meluch, *Selection and Evaluation of Membrane Materials for Liquid Separations*, American Chemical Society, Washington, D.C., 1985.
14. R. Rautenbach and R. Albrecht, *Membrane Processes*, John Wiley & Sons, Inc., New York, 1989.
15. C. Fell, in R. Noble and S. Stern, eds., *Membrane Separation Technology*, Elsevier Science, Inc., New York, 1995, pp. 113–140.
16. R. Kesting, *Synthetic Polymeric Membranes: A Structural Perspective*, Wiley-Interscience, New York, 1985.
17. I. Cabasso, in J. I. Kroschwitz, ed., *Encyclopedia of Polymer Science and Engineering*, 2nd ed., Vol. 9, John Wiley & Sons, Inc., New York, 1987, pp. 509–579.
18. W. Koros and co-workers, *Prog. Polym. Sci.* **13**, 339 (1988).
19. R. Baker, in Ref. 10.
20. H. Strathmann, in M. Porter, ed., *Handbook of Industrial Membrane Technology*, Noyes Publications, Park Ridge, N.J., 1990, pp. 1–60.
21. R. Petersen and J. Cadotte, in Ref. 20, pp. 307–348.
22. P. Stanford, *Proceedings of North American Membrane Society Conference*, Portland, Oreg., 1995.
23. T. Peters, *Desalination*, **83**, 159 (1991).
24. V. Gekas, *Desalination*, **68**, 77 (1988).
25. W. Koros, *Chem. Eng. Prog.* **91**, 68 (Oct. 1995).
26. M. Soltanieh and W. Gill, *Chem. Eng. Comm.* **12**, 279 (1981).
27. M. Mazid, *Sep. Sci. Technol.* **19**, 357 (1984).
28. W. Pusch, *Desalination*, **59**, 105 (1986).
29. J. Dickson, in Ref. 11, pp. 1–51.
30. H. Lonsdale, U. Merten, and R. Riley, *J. Appl. Polym. Sci.* **9**, 1341 (1965).
31. H. Burghoff, K. Lee, and W. Pusch, *J. Appl. Polym. Sci.* **25**, 323 (1980).
32. T. Matsuura and S. Sourirajan, *Ind. Eng. Chem. Proc. Des. Dev.* **20**, 273 (1981).
33. H. Mehdizadeh and J. Dickson, *J. Membrane Sci.* **42**, 119 (1989).
34. H. Mehdizadeh, "Modeling of Transport Phenomena in Reverse Osmosis Membranes," dissertation, McMaster University, Hamilton, Ont., Canada, 1990.
35. H. Mehdizadeh and J. Dickson, *Comp. Chem. Eng.* **14**, 157 (1990).
36. D. Bhattacharyya and co-workers, *Chem. Eng. Comm.* **42**, 111 (1986).
37. J. Siler, "Reverse Osmosis Membranes-Concentration Polarization and Surface Fouling: Predictive Models and Experimental Verifications," dissertation, University of Kentucky, Lexington, Ky., 1987.

38. E. Matthiasson and B. Sivik, *Desalination*, **35**, 59 (1980).
39. V. Gekas and B. Hallstrom, *J. Membrane Sci.* **30**, 153 (1987).
40. D. Bhattacharyya, S. Back, and R. Kermode, *J. Membrane Sci.* **48**, 231 (1990).
41. P. Brian, in U. Merten, ed., *Desalination by Reverse Osmosis*, MIT Press, Cambridge, Mass., 1966, pp. 161–202.
42. M. Williams, "Measurement and Mathematical Description of Separation Characteristics of Hazardous Organic Compounds with Reverse Osmosis Membranes," dissertation, University of Kentucky, Lexington, Ky., 1993.
43. H. Ridgway, in Ref. 11, pp. 429–481.
44. G. Belfort, R. Davis, and A. Zydney, *J. Membrane Sci.* **95**, 1–58 (1994).
45. H. Ridgway, M. Rigby, and D. Argo, *J. AWWA*, **77**, 97–106 (1985).
46. J. Lepore and R. Ahlert, in Ref. 11, pp. 141–183.
47. T. Osta and L. Bakheet, *Desalination*, **63**, 71–80 (1987).
48. S. N. Gaeta and co-workers, *Desalination*, **83**, 383 (1991).
49. K. Nita and co-workers, *Desalination*, **96**, 33–41 (1994).
50. R. Singh, *Desalination*, **95**, 27–37 (1994).
51. J. Glater, S.-K. Hong, and M. Elimelech, *Desalination*, **95**, 325–345 (1994).
52. G. Congjie, L. Xueren, and B. Zhiguo, *Desalination*, **83**, 271–278 (1991).
53. A. Ko and D. Guy, in Ref. 11, pp. 185–278.
54. S. Ebrahim, *Desalination*, **96**, 225–238 (1994).
55. S. Ebrahim and H. El-Dessouky, *Desalination*, **99**, 169–188 (1994).
56. A. Jaffer, *Desalination*, **96**, 71–79 (1994).
57. H. Mehdizadeh and J. Dickson, *AIChE J.* **39**, 434–445 (1993).
58. J. Jiang and co-workers, *Desalination*, **17**, 107 (1989).
59. S. Gao and Q. Bao, *J. Liq. Chromatog.* **12**, 2083 (1989).
60. J. Jiang and C. Jiayan, *Desalination*, **78**, 389 (1990).
61. R. Cheng and co-workers, *Desalination*, **85**, 33 (1991).
62. D. Bhattacharyya and co-workers, *EPA Report*, EPA 600/87/053, U.S. Environmental Protection Agency, Washington, D.C., 1987.
63. D. Bhattacharyya and M. Madadi, *AIChE Symposium Series* **84**(261), 139 (1988).
64. D. Bhattacharyya and A. Kothari, *EPA Report*, Cooperative Agreement No. CR814491, U.S. EPA, Washington, D.C., 1992.
65. S. Sourirajan, *Reverse Osmosis*, Academic Press, Inc., New York, 1970.
66. M. Kurihara and co-workers, *Desalination*, **38**, 449 (1981).
67. K. Koyama and co-workers, *J. Appl. Polym. Sci.* **27**, 2845 (1982).
68. K. Koyama and co-workers, *J. Appl. Polym. Sci.* **29**, 2929 (1984).
69. S. Lynch and co-workers, *EPA Report*, EPA 600/1-84/018, U.S. EPA, Washington, D.C., 1984.
70. J. Rickabaugh and co-workers, *Proceedings of the 12th Annual Hazardous Wastes Symposium*, Cincinnati, Ohio, 1986.
71. W. Pusch, Y. Yu, and L. Zheng, *Desalination*, **75**, 3 (1989).
72. R. Rautenbach and A. Grönschl, *Proceedings of the 203rd American Chemical Society National Meeting*, San Francisco, Calif., 1992.
73. E. Lohman, *Desalination*, **96**, 349–358 (1994).
74. Y. Ayyash and co-workers, *Desalination*, **96**, 215–224 (1994).
75. P. Cartwright, *Desalination*, **56**, 17 (1985).
76. K. Linde, A.-S. Jönsson, and R. Wimmerstedt, *Desalination*, **101**, 21–30 (1995).
77. C. Schmidt, I. White, and R. Ludwig, *EPA Report*, EPA 600/R-93/150, U.S. EPA, Washington, D.C., 1993.
78. K. Imasu, *Desalination*, **56**, 137 (1985).
79. G. Hsiue and co-workers, *Desalination*, **71**, 35 (1989).
80. M. Chu, C. Tung, and M. Shieh, *Sep. Sci. Technol.* **25**, 1339 (1990).
81. S. Prabhakar and co-workers, *Sep. Sci. Technol.* **27**, 349 (1992).
82. S. Prabhakar and co-workers, *Sep. Sci. Technol.* **29**(8), 1001–1010 (1994).
83. R. Rautenbach and R. Mellis, *Desalination*, **101**, 105–113 (1995).
84. H. Tsuge and K. Mori, *Desalination*, **23**, 123 (1977).
85. I. Nusbaum and D. Argo, in Ref. 12, p. 377.
86. M. Reinhard and co-workers, *J. AWWA* **78**(4), 163 (Apr. 1986).
87. Y. Suzuki and T. Minami, *Wat. Sci. Tech.* **23**, 1629 (1991).
88. C. Frith, in Ref. 11, pp. 279–310.
89. D. Ulietu, *Proceedings of the 19th Annual Technical Meeting of the Institute of Environmental Science*, Las Vegas, Nev., 1993.
90. Y. Egozy, J. Denoncourt, and G. Ganzi, in Ref. 11, pp. 311–346.
91. P. Parise, B. Parekh, and R. Smith, in Ref. 11, pp. 347–398.
92. P. Sinisgalli and J. McNutt, *J. AWWA* **78**(5), 47 (May 1986).
93. P. Cartwright, *Desalination*, **83**, 225 (1991).
94. C. Slater, R. Ahlert, and C. Uchirin, *Desalination*, **48**, 171 (1983).
95. A. Ghabris, M. Abdel-Jawad, and G. Aly, *Desalination*, **75**, 213 (1989).
96. M. El-Halwagi, *AIChE J.* **38**(8), 1185–1198 (1992).
97. J. Schoeman and co-workers, *Wat. Sci. Tech.* **25**, 79–93 (1992).
98. C. Slater, J. Zielinski, and R. Wendel, *J. Environ. Sci. Health*, **A27**(5), 1175–1193 (1992).
99. F. Awadalla, C. Striez, and K. Lamb, *Sep. Sci. Technol.* **29**, 483–495 (1994).
100. A. Kulkarni, D. Mukherjee, and W. Gill, *Chem. Eng. Comm.* **129**, 53 (1994).
101. T. Thorsen, *Desalination*, **53**, 217 (1985).
102. E. Chian and F. De Walle, *EPA Report*, EPA-600/2-77-186b (1977).
103. C. Slater, R. Ahlert, and C. Uchirin, *Env. Prog.* **2**, 251 (1983).
104. R. Kinman and D. Nutini, *Proceedings of the 1990 International Congress on Membranes and Membrane Processes*, Chicago, Ill., 1990.
105. J. Lepore and R. Ahlert, *Waste Manage.* **11**, 27 (1991).
106. AWWA Membrane Technology Research Committee, *J. AWWA* **84**(1), 59 (Jan. 1992).
107. T. Marhaba and S. Medler, *Proceedings of 1994 National Conference on Environmental Engineering*, 1994.
108. L. Tan and R. Sudak, *AWWA J.* **84**, 79–87 (1992).
109. T. Eisenberg and E. Middlebrooks, *Reverse Osmosis Treatment of Drinking Water*, Butterworth & Co., Boston, Mass., 1986.
110. J. Jacangelo and co-workers, *J. AWWA*, **87**, 64–77 (1995).
111. L. Awerbuch and co-workers, *Desalination*, **76**, 189–197 (1989).
112. T. Ikeda, H. Muragishi, and T. Uemura, *Desalination*, **98**, 391–400 (1994).
113. M. Nyström, L. Kaipia, and S. Luque, *J. Membrane Sci.* **98**, 249 (1995).

114. P. Eriksson, *Env. Prog.* **7**, 58 (1988).
115. J. Cadotte and co-workers, *Desalination*, **70**, 77 (1988).
116. L. Raman, M. Cheryan, and N. Rajagopalan, *Chem. Eng. Prog.* **90**(3), 68–74 (1994).
117. A. Bindoff and co-workers, *Desalination*, **67**, 453 (1987).
118. K. Ikeda and co-workers, *Desalination*, **68**, 109 (1988).
119. M. Afonso and co-workers, *Water Res.* **26**, 1639 (1992).
120. M. Simpson, C. Kerr, and C. Buckley, *Desalination*, **64**, 305 (1987).
121. M. Perry and C. Linder, *Desalination*, **71**, 233 (1989).
122. Anon., *Food Eng.* **60**, 124 (1988).
123. P. Fu and co-workers, *J. AWWA* **86**(12), 55–72 (1994).
124. K. Agbekdo and co-workers, *Rev. Sci. Eau*, **7**(2), 183–200 (1994).
125. R. Rautenbach and R. Mellis, *Desalination*, **95**, 171–188 (1994).
126. D. Bhattacharyya, R. Adams, and M. Williams, in D. Butterfield, ed., *Biological and Synthetic Membranes*, Alan R. Liss, New York, 1989.
127. M. Williams, R. Deshmukh, and D. Bhattacharyya, *Env. Prog.* **9**, 118 (1990).
128. D. Bhattacharyya and M. Williams, *EPA Report*, EPA 600/2-91/045 (1992).
129. C. Dyke and C. Bartels, *Env. Prog.* **9**, 183 (1990).
130. C. Brouckaert and C. Buckley, *Water SA*, **18**(3), 215–224 (1992).
131. A. Malek, M. Hawlader, and J. Ho, *Desalination*, **99**, 19–38 (1994).
132. H. Niemi and S. Palosaari, *J. Membrane Sci.* **84**, 123–137 (1993).
133. S. Talati, *Desalination*, **97**, 353–361 (1994).
134. R. Ray, in Ref. 9, pp. 355–390.
135. J. Birkett, in Ref. 11, pp. 484–504.
136. S. Ebrahim and M. Abdel-Jawad, *Desalination*, **99**, 39–55 (1994).

D. Bhattacharyya
W. C. Mangum
University of Kentucky

M. E. Williams
EET Corporation

REYNOLDS NUMBER

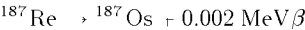
. See FLUIDIZATION; FLUID MECHANICS; RHEOLOGICAL MEASUREMENTS; SEDIMENTATION

RHENIUM AND RHENIUM COMPOUNDS

Rhenium

Rhenium [7440-15-5] is the 75th element in the Periodic Table and the heaviest element in Group 7 (VIIB). Its congeners in Group 7 are manganese and technetium, elements 25 and 43, respectively. The existence of rhenium, predicted in the mid-1800s by Mendeleev, was verified in 1925 by x-ray spectral analysis of samples of columbite, gadolinite, molybdenite, and platinum ores (1). The element was named after the German province of Rhineland. Several papers claiming the identification of this element arose from other laboratories following the initial report (2).

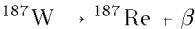
Rhenium, atomic wt 186.2, occurs in nature as two nuclides: ¹⁸⁵Re [14391-28-7], mass 184.9530, in 37.500^o abundance; and ¹⁸⁷Re [14391-29-8], mass 186.9560, in 62.500^o abundance. The latter isotope is radioactive, emitting very low energy radiation and having a half-life estimated at 4.3 (±0.5) × 10¹⁰ yr. The radioactive decay of this isotope has been used to date accurately the time of Earth's formation.



In addition to ¹⁸⁵Re and ¹⁸⁷Re, 16 other radioactive rhenium isotopes are known; none has an appreciable half-life.

Radioactivity associated with ¹⁸⁷Re can be detected only by using sophisticated laboratory equipment because of the low energy of the emitted β-particles. This radioactivity poses no health or safety hazards. Samples of the metal and its compounds are not labeled as radioactive, and typical precautions associated with radioactive materials are not taken during use and handling of the element or its compounds.

Occurrence and Recovery. Rhenium is one of the least abundant of the naturally occurring elements. Various estimates of its abundance in Earth's crust have been made. The most widely quoted figure is 0.027 atoms per 10⁶ atoms of silicon (0.05 ppm by wt) (3). However, this number, based on analyses for the most common rocks, ie, granites and basalts, has a high uncertainty. The abundance of rhenium in stony meteorites has been found to be approximately the same value. An average abundance in siderites is 0.5 ppm. In lunar materials, ¹⁸⁷Re, when compared to ¹⁸⁵Re, appears to be enriched by 1.4^o to as much as 29^o, relative to the terrestrial abundance. This may result from a nuclear reaction sequence beginning with neutron capture by tungsten-186, followed by β-decay of ¹⁸⁷W of a half-life of 24 h (4) (see EXTRATERRESTRIAL MATERIALS).



Rhenium is present in seawater as perthenate ion in very low concentrations. One rhenium mineral, dzhezkazganite, is known; it is found associated with copper deposits in the former Soviet Union. The mineral is believed to be a solid solution of MoS₂ and ReS₂. Rhenium is also found in trace amounts in a variety of other minerals, including gadolinites, alvites, and columbites; in certain manganese, platinum, and uranium ores; and in certain uraniferous ores and vandiferous shales, although none contains rhenium in high enough concentration to make commercial recovery feasible. In uranium and ilsemaninite ores from the Colorado plateau, rhenium occurs as (Re₂O₇)_x [1314-68-7]. Otherwise, rhenium is present as a sulfide.

Occurrence of rhenium and molybdenum together in nature is a consequence of the similarities of these elements. Both elements have a high affinity for sulfide ion. Moreover, the radii of Re⁴⁺ and Mo⁴⁺, 0.74 nm and 0.70 nm, respectively, are almost identical, so that ReS₂ [12038-63-0] and MoS₂ have similar crystal structures with almost identical dimensions (see MOLYBDENUM COMPOUNDS).

As of the 1990s, rhenium is obtained from molybdenite concentrates obtained from porphyry copper (qv) ores. Rhenium, present in small amounts in these ores, is concentrated in the molybdenite by-products to values as high as 2000 ppm, making recovery realistic. Rhenium supplies come from porphyry copper mines in Canada (Island Copper Mine of Utah International, Inc.), in Chile (Chuquicamata, El Teniente, and El Salvador mines, owned by the Chilean government and managed by a state agency, the Corporación Nacional de Cobre de Chile, CODELCO), in Tajikistan and Uzbekistan in central Asia, in Peru, and in the United States (in Utah, Arizona, Nevada, and New Mexico). As of the mid-1990s, additional sources of rhenium are being developed in Mexico and in the former Yugoslavia. A copper smelting operation in mainland China involving one of the world's largest copper deposits began operation in 1984. This deposit also contains recoverable rhenium.

The Chilean deposits have an average grade of rhenium ranging from 230 to 570 ppm. The El Teniente mine is said to hold the world's largest reserves of rhenium, an estimated 680 metric tons. This ore also contains 1.5 wt % Cu and 0.04 wt % Mo.

Most of the processing of the molybdenite by-product to obtain rhenium is carried out in the United States and Germany. Some Canadian concentrate is processed in the United States and returned to the owners in Canada for resale.

The traditional route by which rhenium is obtained involves roasting of the molybdenite concentrate, which converts most of the rhenium to Re₂O₇. This oxide is volatile and exits the system along with some MoO₃ and sulfur oxides. These water-soluble compounds are concentrated in a wet-scrubbing system to which sodium hydroxide is added to react with the sulfur oxides. Processing involving anion exchange or solvent extraction permits separation of ReO₄ from other materials. Treatment of solutions of perthenate ion using HCl and hydrogen sulfide gas produces ReS₂ as black mud, which is separated. Addition of H₂O₂ and NH₄OH produces a solution of NH₄ReO₄ [13598-65-7]. This salt, which crystallizes upon evaporation of the solvent, is either sold directly or converted to the metal by reduction with hydrogen gas.

In earlier procedures, the ReO₄⁻ anion was precipitated from water as the relatively insoluble potassium salt. Reduction of KReO₄ with hydrogen gas gives rhenium metal, but the metal is contaminated with ca 0.4 wt % potassium that cannot be separated easily. Although suitable for some purposes, rhenium formed from KReO₄ is found to be unsatisfactory in applications such as those for use in filaments in mass spectrometer systems. The route involving NH₄ReO₄ avoids this problem.

Before scrubbing procedures were established for copper ore, most of the rhenium was lost as the volatile (Re₂O₇)_x. A small portion, perhaps 10^o %, was retained in flue dust, which was processed to give the metal. A commercial flotation (qv) process for the recovery of the molybdenite by-product is available that permits a high recovery of molybdenum and rhenium. This process is used at the Caridad copper mine in Mexico.

Physical Properties

Selected physical properties of rhenium are summarized in Table 1. The metal is silvery-white and has a metallic luster. It has a high density (21.02 g cm³). Only platinum, iridium, and osmium have higher densities. The melting point of rhenium is higher than that of all other elements except tungsten (mp 3410°C) and carbon (mp 3550°C).

Table 1. Selected Physical Properties of Rhenium

Property	Value
atomic number	75
atomic weight	186.2
mp, °C	3180
bp, °C	5926 ^a
density, g cm ³	21.02
crystal structure	hexagonal close-packed
lattice constants, nm	
<i>a</i>	0.2760
<i>c</i>	0.4458
metal radius, 12 coordinate, nm	0.1372
ΔH_{subl} , kJ mol ^b	791
<i>S</i> ^o , crystal, J (mol·K) ^b	37.2 ± 1.2
specific heat, 20°C, J (mol·K) ^b	25.1
ionization potentials, kJ ^b	
Re → Re ⁺ + <i>e</i>	757
Re ⁺ → Re ²⁺ + <i>e</i>	1597
Re ²⁺ → Re ³⁺ + <i>e</i>	2502
electrical resistivity, 20°C, μΩ·cm	19.3
thermal coefficient of electrical resistivity, °C ^{−1}	3.95 × 10 ^{−3}
Richardson constant, kJ ^b	462 ^c
thermal conductivity, 0–100°C, W (cm·K)	0.400
module of elasticity, Pa ^d	0.46

^a Estimated value.
^b To convert J to cal, divide by 4.184.
^c This value is equivalent to 4.8 eV.
^d To convert Pa to mm Hg, multiply by 0.0075.

Rhenium metal exhibits a very low paramagnetism which is field-independent and almost temperature-independent. In the range 79–471 K, magnetic susceptibility data for two independent samples were shown to correspond, when *T* is in Kelvin, to the following equations (5):

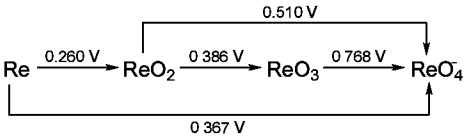
$$\chi_g \times 10^6 = 0.280 + (7.5 \times 10^{-6}) T$$

$$\chi_g \times 10^6 = 0.310 + (17 \times 10^{-6}) T$$

Chemical Properties

Rhenium does not react with atmospheric oxygen at ambient temperatures, but at higher temperatures it burns, giving (Re₂O₇)_x. This volatile yellow crystalline compound dissolves in water to give perrhenic acid [13768-11-1], HReO₄. Reactions of the metal with chlorine or bromine produce Re₂Cl₁₀ [39368-69-9] and Re₂Br₁₀ [30937-53-2]; reaction with fluorine gives ReF₆ [14009-17-9] and ReF₇ [17029-21-9]. Rhenium is not affected by water and hydrohalic acids but reacts quickly with HNO₃ to produce HReO₄. Fusing the metal with NaOH and oxidizing agents such as KNO₃ or Na₂O₂ produces perrrhenate salts. Aqueous 30 wt % H₂O₂ oxidizes the metal to HReO₄. However, the rate of this reaction is dependent on the nature of the sample.

The oxidation potentials for rhenium in aqueous acidic solution are summarized in the following diagram (6).



The relatively low potential for the oxidation of rhenium to perrrhenate ion accounts for the perrrhenate ion being a much weaker oxidizing agent than permanganate. This accounts for ReO₄ formation in most oxidizing systems and its inability to serve as a strong oxidizing agent. The absence of aqueous Re²⁺ chemistry also contrasts with the chemistry of manganese.

Volumetric, gravimetric, and spectrophotometric methods for the analysis of rhenium are mainly of historical interest. These are described elsewhere (7).

Alloys

A significant property of rhenium is its ability to alloy with molybdenum and tungsten. Phase diagrams for Mo–Re and W–Re are given in Figures 1 and 2, respectively (8). Molybdenum alloys containing up to 50 wt % Re are body-centered cubic (bcc) solid solutions. An abrupt increase in ductility occurs just below the solubility limit, at ca 47 wt % Re. An alloy having 50 wt % Re can be fabricated by either warm or cold working. Unlike molybdenum, this alloy is ductile, even at temperatures below −196°C. The alloy can be welded. The addition of rhenium to tungsten gives alloys of improved properties. The alloy of composition 25 wt % Re is a bcc solution. This alloy has improved ductility and a lower ductile-to-brittle transition temperature than pure tungsten.

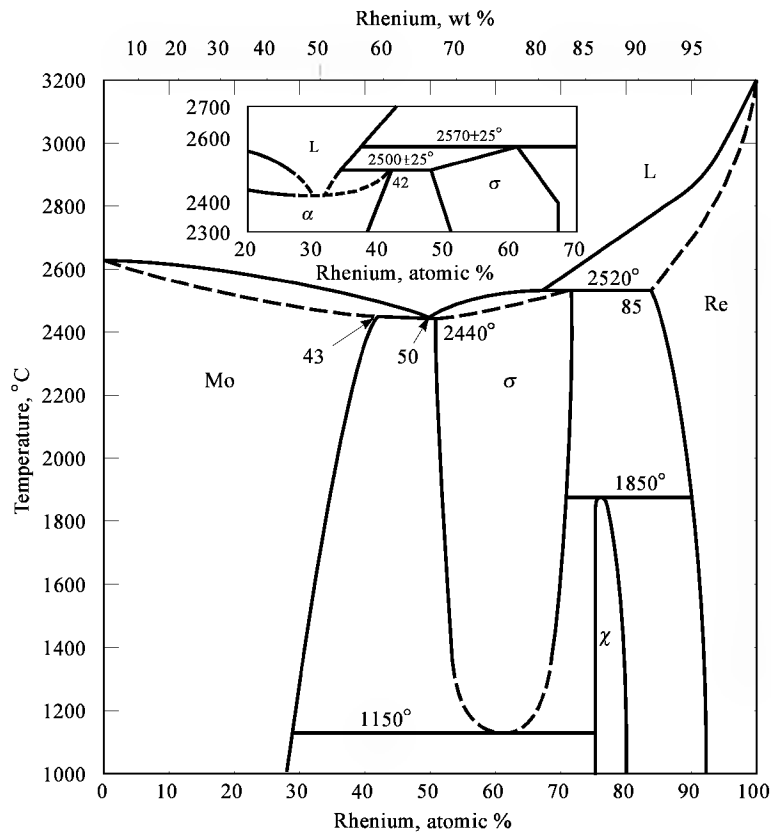


Fig. 1. The Mo–Re alloy system, where the insert shows the more complicated midrange (6). Temperatures (values with °) are given on lines in centigrade. Other values refer to atomic % Re.

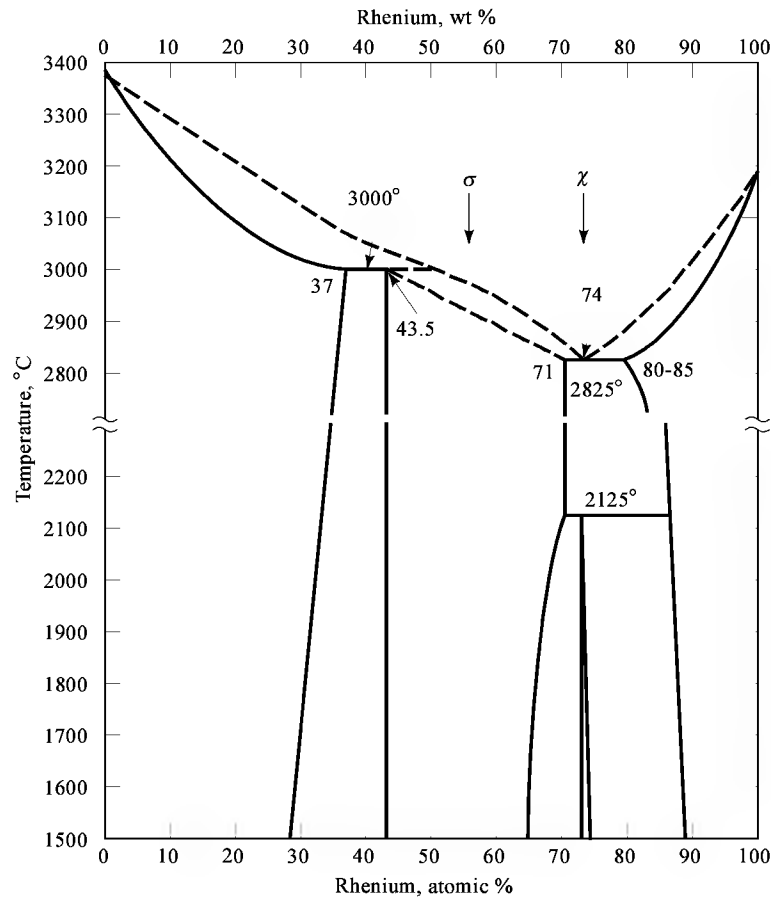


Fig. 2. The W–Re alloy system (6). Temperature (values with °) in centigrade; other values = at. % Re.

The addition of thorium to a Re–W alloy produces a material having 74 wt % W, 24 wt % Re, and 2% ThO₂ that is used for heated cathodes in electron tubes. This material has good ductility and high resistance to breaking by mechanical shock.

Nickel–rhenium alloys containing thorium or other additives have been developed for use as cathodes on electrovacuum devices. Rhenium is found to improve the strength properties substantially. At 1000°C the strength of a 90 wt % Ni–10% Re alloy exceeds the strength of a Ni–V cathode material by 90%. Rigidity is exceeded by a factor of 150 to 200%.

Metal and Alloy Processing. Rhenium and alloys of rhenium with molybdenum and tungsten can be consolidated by powder metallurgy (see METALLURGY, POWDER METALLURGY). The process can be carried out using blended powders but better results arise using pre-alloyed material. After screening, sizing, and blending of the powder, bars are pressed either mechanically or isostatically at pressures of 210–340 MPa (30,000–50,000 psi). This material requires sintering at 1200°C in hydrogen or under vacuum. For W–Re alloys where diffusion rates are slower, the sintering temperature should be

as high as possible to avoid formation of the rhenium-rich σ -phase that is brittle. Use of pre-alloyed particles permits sintering at temperatures as low as 2200°C.

The Mo–Re and W–Re alloys can also be produced by electron beam or arc melting. However the resulting grain size is much larger so fabrication by forging or extrusion is desirable. The principal advantage of melting is that homogeneity is assured.

Fabrication of pure rhenium is done cold with recrystallization annealing after each 10–40% reduction, depending on the operation and size. Annealing of Mo–Re and W–Re products is also necessary.

Uses. Rhenium finds use as filaments in electron tubes, light bulbs, and in photoflash bulbs. It is particularly useful for filaments used in mass spectrometers. The high melting point and low vapor pressure of rhenium are prerequisites to this high temperature usage. Alloys of molybdenum or tungsten are used as heating elements, but this use requires either evacuated systems or an inert atmosphere because rhenium is oxidized by atmospheric oxygen at high temperatures. About 8% of the rhenium produced is used in these applications.

The electrical resistance of rhenium is about 3.5 times higher than tungsten at 20°C. However, the difference is reduced at higher temperatures because of rhenium's lower temperature coefficient, so that at 2500°C the resistivity of rhenium is only about 20% more than tungsten. The higher resistance at low temperature, combined with a lower temperature coefficient, contributes to rapid heating of a filament.

Rhenium exhibits a greater resistance than tungsten to the water cycle effect, in which lamps and electron tubes become blackened by deposition of metal. This phenomenon involves catalysis by small quantities of water that react with the metal in a hot filament to produce a volatile metal oxide and hydrogen. The oxide condenses on the surface of the bulb and is reduced back to the metal by hydrogen.

Rhenium may be electroplated from solutions of ReO_4 in sulfuric acid. To obtain a durable rhenium plate it is desirable to deposit the metal in thin layers, annealing at high temperature after each layer is applied.

The thermionic work function for rhenium (4.8 eV) is slightly higher than the value for tungsten but still low enough so that this metal can be used in thermionic devices. Alloys containing thorium and tungsten or nickel are often used in this context. These alloys have a higher emission, improved ductility, and greater shock resistance (see THERMOELECTRIC ENERGY CONSERVATION).

Two alloys containing tungsten are commercially available. The first, containing about 3 wt % rhenium, is used for heating filaments. The rhenium contributes improved resistance to thermal and mechanical shock. The second alloy contains about 25 wt % rhenium. This latter alloy is sold as sheet, rod, and heavy wire and may be fabricated for various uses. An important use of these rhenium alloys is in the construction of thermocouples. Various combinations, 3 wt % Re–97 wt % W, or 25 wt % Re–75 wt % W, are useful for measurement of temperatures to 2500°C (see TEMPERATURE MEASUREMENT).

Rhenium coatings are used on some electrical contacts because of the metal's resistance to arc erosion wear. Rhenium is also used in contacts for engine magnetos. On interrupting the current, a thin oxide coating is formed that prevents the contacts from sticking or being welded together. The oxide coating is also conducting and does not impair further efficient operation.

A rhenium–molybdenum alloy is reported to be superconducting at 10 K.

Rhenium-Containing Catalysts. As of the mid-1990s, the largest use of rhenium by far is in bimetallic petroleum (qv) reforming catalysts used in the production of unleaded and low lead gasoline (see GASOLINE AND OTHER MOTOR FUELS). These catalysts contain approximately 0.3 wt % platinum and 0.3 wt % rhenium. In newer catalyst systems a somewhat higher proportion of rhenium is used. The bimetallic systems have an advantage over monometallic systems which usually contain platinum alone, because the former reduce the aging rate of the catalysts. Platinum–rhenium catalysts are also used in the production of benzene, toluene, and xylenes by reforming (see BTX PROCESSING). The use of rhenium catalysts for the liquid-phase reduction of nitrobenzenes has been studied. Such catalysts appear to be comparable to palladium systems and much superior to nickel. Rhenium catalysts have also been used to carry out olefin metathesis reactions.

Various studies have been carried out to understand the nature and function of rhenium catalysts (9). Exafs spectroscopy has been used to determine the chemical composition of various rhenium species on the surface of a MgO support. A catalyst formed by deposition of single rhenium atoms on the support surface was active for olefin hydrogenation but not cyclopropane hydrogenolysis; whereas a catalyst having Re_3 units serves both functions. These Re_3 catalysts are made by deposition of rhenium cluster complexes on the support. The use of rhenium on $\gamma\text{-Al}_2\text{O}_3$ may be favored over the use of Groups 8–10 (VIII B) metals because rhenium has a higher affinity toward the oxygen atoms on the support surface.

Rhenium Compounds

As a general rule, elements in the second and third transition series have similar chemical properties. In contrast, the properties of the first member of the series are often different. This pattern of behavior is seen in Group 7 (VIIB). The properties of rhenium and technetium differ considerably from those of manganese.

Compounds of rhenium having metal oxidation states between -1 and $+7$ are known. As expected, the lower oxidation states (-1 , 0 , $+1$) are encountered only with carbonyl groups and other π -acceptor ligands, and in these oxidation states all elements of this periodic group have fairly similar behavior. There is a significant difference between manganese and rhenium, however. Many more rhenium compounds than manganese compounds having oxidation states of $+4$, $+5$, $+6$, and $+7$ are known. Although rhenium(II) compounds are known to exist as solids and in nonaqueous solution, no aqueous chemistry of the Re^{2+} ion is known. The $+2$ oxidation state is common for manganese.

A particularly significant part of rhenium chemistry involves cluster compounds in which there is metal–metal bonding. This chemistry centers largely around the $+3$ oxidation state.

A survey of known types of rhenium compounds is presented in Table 2. The selected examples include a number of commonly encountered compounds and emphasize the diversity in structures.

Table 2. Examples of Rhenium Compounds

Oxidation state	Electronic configuration	Coordination number	Metal atom geometry	Example ^a	CAS Registry Number
I	d^8	5	trigonal bipyramid	$\text{Na}(\text{Re}(\text{CO})_5)$	[33634-75-2]
0	d^7	6	octahedral	$\text{Re}_2(\text{CO})_{10}$	[14285-68-8]
I	d^6		octahedral	$\text{ReCl}(\text{CO})_5$	[14099-01-5]
				$\text{K}_5[\text{Re}(\text{CN})_6]$	[39700-07-7]
				$(\text{Re}(\text{CNC}_5\text{H}_4)_2)\text{I}$	[81195-37-1]
				$\text{Re}(\text{CO})_3(\eta\text{-C}_5\text{H}_5)_2$	[12079-73-1]
				$\text{ReCl}(\text{dppe})_2(\text{N}_2)$	[25349-29-5]
II	d^5	6	octahedral	$\text{ReCl}_2(\text{dppe})_2$	[109428-32-2]
III	d^4	6	octahedral	$\text{ReCl}_3(\text{P}(\text{CH}_3)_2\text{C}_6\text{H}_5)_3$	[15613-32-8]
			trigonal prism	$\text{Re}(\text{S}_2\text{C}_2(\text{C}_6\text{H}_5)_2)_3$	[14264-08-5]
		7		$\text{K}_4(\text{Re}(\text{CN})_7) \cdot 2\text{H}_2\text{O}$	[59370-53-5]
			square pyramid	$\text{K}_2(\text{Re}_2\text{Cl}_8)$	[13841-78-6]
		7	capped octahedral	$\text{ReH}_3(\text{dppe})_2$	[17300-50-4]
IV	d^3	6	octahedral	K_2ReCl_6	[16940-97-9]

V	d^2	6	octahedral	ReCl ₄	[13569-71-6]
				K ₄ (Re ₂ OCl ₁₀)	[19979-02-3]
				Re ₂ Cl ₁₀	[39368-69-9]
				ReOCl ₃ (P(C H ₅) ₃) ₂	[17442-18-1]
VI	d^1	8	square pyramid	K ₂ (ReOCl ₅)	[17443-52-6]
				ReH ₅ (P(C H ₅)(C ₂ H ₅) ₂) ₃	[12104-30-2]
		5	square pyramid	ReOCl ₄	[13814-76-1]
				ReF	[10049-17-9]
VII	d^0	6	octahedral	Re(CH ₃)	[56090-02-9]
				K ₂ ReF ₈	[57300-90-0]
		8	square antiprism	KReO ₄	[10466-65-6]
				ReO ₅ Cl	[7791-09-5]
		4	tetrahedral	(Re ₂ O ₇) _x	[1314-68-7]
				ReF ₇	[17029-21-9]
		9	tricapped trigonal prism	K ₂ ReH ₉	[25396-46-7]

^a dppe = 1,2-bis(diphenylphosphino)ethane.

Rhenium Carbonyls and Related Compounds. The parent compound of the low valent rhenium compounds is Re₂(CO)₁₀. Dirhenium decacarbonyl [14285-68-8], a white crystalline compound, mp 177°C, is volatile and soluble in most organic solvents. Its preparation in a high pressure reaction between Re₂O₇, H₂, and CO was reported in 1941. It has a molecular structure of two square pyramidal Re(CO)₅ groups linked by a metal–metal bond. This compound is available commercially as a specialty chemical. It is the precursor to other low valent rhenium carbonyl compounds, including the halides, ReX(CO)₅, where X = Cl, Br, or I; alkyl, aryl, and acyl compounds, Re(R)(CO)₅; the hydride complexes ReH(CO)₅ [16457-30-0], Re₂(μ-H)₂(CO)₈ [38887-05-7], and Re₃(μ-H)₃(CO)₁₂ [12146-47-3]; and hydrocarbon complexes, including Re(CO)₃(η-C₅H₅) and [Re(NO)(CO)₂(η-C₅H₅)]PF₆ [12306-73-9]. Research in the 1980s and 1990s has focused on the photochemical activation of H₂ by Re₂(CO)₁₀, formation of metal atom clusters, and on the reduction of a coordinated carbon monoxide ligand to methane or methanol. The latter reaction has received much attention because it provides mechanistic information about the Fischer-Tropsch reaction (see CARBONYL; COAL CONVERSION PROCESSES).

Pyrolysis of Re₂(CO)₁₀ at 400°C *in vacuo* or in an inert atmosphere has been used to obtain pure rhenium metal.

Oxides and Sulfides. Rhenium reacts with O₂ at moderate temperatures to give the yellow solid (Re₂O₇)_x, mp 220°C. This material contains tetrahedral and octahedral rhenium atoms in a lattice structure, so arranged that the formation of molecular Re₂O₇ can readily occur upon vaporization. Pertrhenic acid, a strong acid, is formed when (Re₂O₇)_x dissolves in water. It can be isolated upon evaporation of this solvent as crystalline Re₂O₇·2H₂O [41017-51-6]. Lower oxides of thenium, such as ReO₃ [1314-28-9] and ReO₂ [12036-09-8], are formed by heating mixtures of (Re₂O₇)_x and Re in the proper mole ratio. The hydrated compound, ReO₂·2H₂O [54706-20-6], is formed by hydrolysis of [ReCl₆]²⁻. A black heptasulfide, Re₂S₇ [12038-67-4], is obtained when a ReO₄ solution in HCl is saturated with H₂S. This may be reduced to ReS₃ [12166-08-4] by H₂ or to ReS₂ upon heating at 750°C under nitrogen. Also, ReS₂ is the product formed by direct reaction of rhenium and sulfur.

Salts of pertrhenic acid may be obtained in acid–base reactions, and may include the tetrahedral ReO₄⁻ anion or the octahedral anion ReO₅⁻, eg, in Ba₅(ReO₅)₂ [13598-09-9]. Ammonium pertrhenate and pertrhenic acid, as well as rhenium metal, are sold by the primary suppliers of this element.

Rhenium oxides have been studied as catalyst materials in oxidation reactions of sulfur dioxide to sulfur trioxide, sulfite to sulfate, and nitrite to nitrate. There has been no commercial development in this area. These compounds have also been used as catalysts for reductions, but appear not to have exceptional properties. Rhenium sulfide catalysts have been used for hydrogenations of organic compounds, including benzene and styrene, and for dehydrogenation of alcohols to give aldehydes (qv) and ketones (qv). The significant property of these catalyst systems is that they are not poisoned by sulfur compounds.

Rhenium Halides and Halide Complexes. Rhenium reacts with chlorine at ca 600°C to produce rhenium pentachloride [39368-69-9], Re₂Cl₁₀, a volatile species that is dimeric via bridging halide groups. Rhenium reacts with elemental bromine in a similar fashion, but the metal is unreactive toward iodine. The compounds ReCl₄, ReBr₄ [36753-03-4], and ReI₄ [59301-47-2] can be prepared by careful evaporation of a solution of HReO₄ and HX. Substantiation in a modern laboratory would be desirable. Lower oxidation state halides (Re₃X₉)_x are also prepared from the pentavalent or tetravalent compounds by thermal decomposition or chemical reduction.

Many coordination complexes of these halide complexes are known. More common examples have the formulas ReX₄L₂, (Re₂X₈)²⁻, Re₃X₉L₃, and (Re₃X₁₂)³⁻, where L indicates a ligand and X is a halide ion. Rhenium(IV) complexes form in reductive reactions between the ligand and Re₂Cl₁₀. For example, reduction occurs spontaneously when Re₂Cl₁₀ dissolves in acetonitrile, giving the product ReCl₄(NCCH₃)₂ [116853-53-5]. The complexes of rhenium(III) are obtained from (Re₃Cl₉)_x, (Re₃Br₉)_x [13569-49-8], and (Re₃I₉)_x [15622-42-11] and added ligand. Rhenium(III) complexes contain metal–metal bonds. The structure of [Re₂Cl₈]²⁻ has two square planar ReCl₄ units linked by a short (0.2237-nm) rhenium–rhenium quadruple bond. In salts of [Re₃X₁₂]³⁻ there is a triangle of rhenium atoms, linked by metal–metal single bonds (ca 0.25 nm) and bridging halide atoms. The same triangular metal structural unit is retained in polymeric (Re₃X₉)_x.

Rhenium Hydrides and Alkyls. Reduction of KReO₄ by potassium in ethanol gives K₂ReH₉ [25396-46-7]. The identity of this compound was unknown until its structure was determined by a neutron diffraction study. The anion contains rhenium and nine hydride ligands in a tricapped trigonal prismatic arrangement. Reactions of this species with various phosphines give (ReH₈L)₃, ReH₇L₂, and ReH₆L₃, where L = tertiary phosphines. The neutral complexes are also formed from phosphine–rhenium halide complexes with LiAlH₄. Reactions of these complexes with unsaturated hydrocarbons have been much studied as models for various catalysts.

Compounds of the formulas Re(CH₃)₅, ReO(CH₃)₄, Li₂[Re₂(CH₃)₈] [60975-25-9], ReO₂(CH₃)₃ [56090-011-8], and ReO₃CH₃ [70197-13-6] have been prepared. The first two compounds were obtained from reaction of rhenium halides or oxyhalides and methylolithium; the last three were formed from the species by oxidation or reduction reactions. The use of these hydride and alkyl complexes as catalysts is under investigation.

Health and Safety Factors

Little is known concerning the toxicology of rhenium or its compounds. Rhenium heptoxide is said to be an irritant if inhaled.

Economic Aspects

World production of rhenium in 1993 and 1994 was estimated to be 34 and 35 metric tons, respectively. The price of rhenium metal in 1995 varied in the range \$825–\$1600 /kg. Price is based on form, purity, and quantity. The price for the most commonly used salt, NH₄ReO₄, was \$550–\$770 /kg. Dirhenium decacarbonyl and pertrhenic acid are also commercially available.

BIBLIOGRAPHY

"Rhenium and Rhenium Compounds" in *ECT* 1st ed., Vol. 11, pp. 721–730, by C. F. Hiskey, Polytechnic Institute of Brooklyn; "Rhenium" in *ECT* 2nd ed., Vol. 17, pp. 411– 423, by J. H. Port, Cleveland Refractory Metals Division, Chase Brass and Copper Co., Inc.; "Rhenium and Rhenium Compounds" in *ECT* 3rd ed., Vol. 20, pp. 249–258, by P. M. Treichel, University of Wisconsin.

1. W. Noddack, I. Tacke, and O. Berg, *Naturwissenschaften* **13**, 567 (1925).
2. M. E. Weeks and H. M. Leicester, *The Discovery of the Elements*, 7th ed., Chemical Education Press, Easton, Pa., 1968, pp. 823–827; J. G. F. Druce, *Rhenium*, Cambridge University Press, Cambridge, U.K., 1948.
3. L. H. Aller, *The Abundance of the Elements*, Vol. 7, Interscience Publishers, Inc., New York, 1961, pp. 32–33; L. H. Ahrens and S. R. Taylor, *Spectrochemical Analysis*, Addison Wesley Press, Reading, Mass., 1961, p. 93; K. Rankama, *Isotope Geology*, Pergamon Press, London, 1954, p. 135.
4. R. Michel, U. Herpers, H. Kulus, and W. Herr, *J. Radioanal. Chem.* **17**, 177 (1973).
5. R. W. Asmussen and H. Soling, *Acta Chem. Scand.* **8**, 563 (1954).
6. J. P. King and J. W. Cobble, *J. Am. Chem. Soc.* **79**, 1550 (1957).
7. R. Colton, *The Chemistry of Rhenium and Technetium*, Wiley-Interscience, New York, 1965, pp. 20–26; R. D. Peacock, *The Chemistry of Technetium and Rhenium*, Elsevier, Amsterdam, the Netherlands, 1966, pp. 122–125.
8. R. P. Elliot, *Constitution of Binary Alloys*, 1st Suppl., McGraw-Hill Book Co., Inc., New York, 1965, pp. 629, 776.
9. B. E. Gates, *Catalytic Chemistry*, John Wiley & Sons, Inc., New York, 1992, pp. 345, 382.

General References

R. D. Peacock in A. F. Trotman-Dickinson, ed., *Comprehensive Inorganic Chemistry*, Vol. 3, Pergamon Press, London, 1973, p. 905.
F. A. Cotton and G. Wilkinson, *Advanced Inorganic Chemistry*, 5th ed., Wiley-Interscience, New York, 1988, p. 847.
L. J. Alverson, "Rhenium," in *Minerals Yearbook*, Vol. 1, Bureau of Mines, U.S. Government Printing Office, Washington, D.C., 1980, p. 743.
D. F. C. Morris and E. L. Short, *Handbook of Geochemistry*, Springer-Verlag, Heidelberg, Germany, 1965, p. 75-1.
R. D. Peacock, *The Chemistry of Technetium and Rhenium*, Elsevier, Amsterdam, the Netherlands, 1966.
R. Colton, *The Chemistry of Rhenium and Technetium*, Wiley-Interscience, New York, 1965.
G. Rouschias, *Chem. Rev.* **74**, 531 (1974).
K. A. Conner and R. A. Walton, *Rhenium*, in G. Wilkinson, R. D. Gillard, and J. A. McCleverty, eds., *Comprehensive Coordination Chemistry*, Vol. 4, Pergamon Press, London, 1987, p. 125.

Paul M. Treichel, Jr.
University of Wisconsin, Madison

RHEOLOGICAL MEASUREMENTS

Rheology is the science of the deformation and flow of matter. It is concerned with the response of materials to applied stress. That response may be irreversible viscous flow, reversible elastic deformation, or a combination of the two. Control of rheology is essential for the manufacture and handling of numerous materials and products, eg, foods, cosmetics, rubber, plastics, paints, inks, and drilling muds. Before control can be achieved, there must be an understanding of rheology and an ability to measure rheological properties.

Deformation is the relative displacement of points of a body. It can be divided into two types: flow and elasticity. Flow is irreversible deformation; when the stress is removed, the material does not revert to its original form. This means that work is converted to heat. Elasticity is reversible deformation; the deformed body recovers its original shape, and the applied work is largely recoverable. Viscoelastic materials show both flow and elasticity. A good example is Silly Putty, which bounces like a rubber ball when dropped, but slowly flows when allowed to stand. Viscoelastic materials provide special challenges in terms of modeling behavior and devising measurement techniques.

The study of flow and elasticity dates to antiquity. Practical rheology existed for centuries before Hooke and Newton proposed the basic laws of elastic response and simple viscous flow, respectively, in the seventeenth century. Further advances in understanding came in the mid-nineteenth century with models for viscous flow in round tubes. The introduction of the first practical rotational viscometer by Couette in 1890 (1,2) was another milestone. In the late twentieth century the science of rheology has grown rapidly.

This article is concerned with rheological measurements on both liquids and solids and the principles on which they are based. The flow properties of a liquid are defined by its resistance to flow, ie, viscosity, and may be measured by determining the rate of flow through a capillary, the resistance to flow when the fluid is sheared between two surfaces, or the rate of motion of a bubble or ball moving through the fluid. The mechanical properties of an elastic solid may be studied by applying a stress and measuring the deformation or strain. Many solids, such as polymers, undergo flow in addition to recoverable elastic deformation. Furthermore, a number of liquids show elastic as well as flow behavior. These materials are viscoelastic, and additional techniques beyond those indicated for solids and liquids are needed for complete characterization. Examples of such methods are the measurement of response to sinusoidal oscillatory motion; the measurement of flow with time after application of stress, ie, creep; and the measurement of the rate and degree of recovery after removal of stress, ie, creep recovery or recoil.

In addition to viscometers, optical devices such as microscopes and cameras can be used for defining and solving flow problems as well as characterizing materials (3–5). Optical techniques allow the investigator to determine the physical structure of the material and visualize its flow processes.

Viscosity

A liquid is a material that continues to deform as long as it is subjected to a tensile and or shear stress. The latter is a force applied tangentially to the material. In a liquid, shear stress produces a sliding of one infinitesimal layer over another, resulting in a stack-of-cards type of flow (Fig. 1).

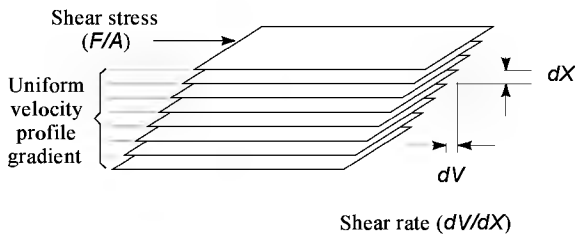


Fig. 1. Laminar flow in simple shear. $F/A = \eta dV/dX$, where F is the force acting on area A , V the velocity and X the thickness of the layer, and η the coefficient of viscosity or the Newtonian viscosity.

For a liquid under shear the rate of deformation or shear rate is a function of the shearing stress. The original exposition of this relationship is Newton's law, which states that the ratio of the stress to the shear rate is a constant, ie, the viscosity. Under Newton's law, viscosity is independent of shear rate. This is true for ideal or Newtonian liquids, but the viscosities of many liquids, particularly a number of those of interest to industry, are not independent of shear rate. These non-Newtonian liquids may be classified according to their viscosity behavior as a function of shear rate. Many exhibit shear thinning, whereas others give shear thickening. Some liquids at rest appear to behave like solids until the shear stress exceeds a certain value, called the yield stress, after which they flow readily.

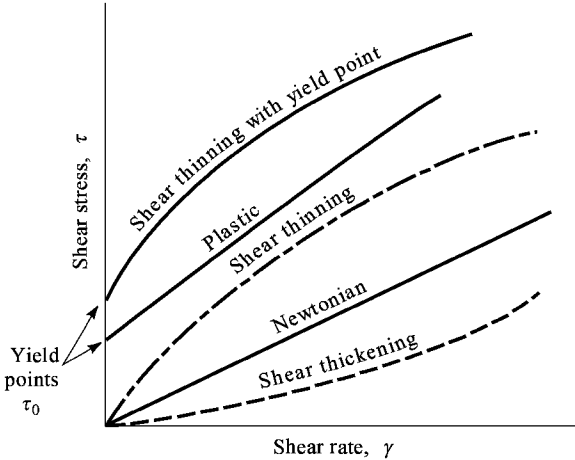


Fig. 2. Flow curves (shear stress vs shear rate) for different types of flow behavior.

Some commonly observed types of flow behavior are shown in Figure 2, in which the shear stress is plotted against shear rate. These plots are called flow curves and are frequently used to express the rheological behavior of liquids. Newtonian flow is shown by a straight line, and shear thinning and thickening by curves. Yield stresses, τ_0 , are shown by intercepts on the stress (y) axis. It should be pointed out that the existence of yield stresses is controversial; they may be artifacts resulting from high Newtonian viscosity at low shear rates (6). However, in many dispersed systems, particularly where severe flocculation occurs, this viscosity is so high that the material would take years to flow. Therefore, in practice, there are true yield stresses. This parameter can be quite useful in characterizing materials. Additional information on yield is available (7–10).

Viscosity is equal to the slope of the flow curve, $\eta = d\tau / d\dot{\gamma}$. The quantity $\tau / \dot{\gamma}$ is the viscosity η for a Newtonian liquid and the apparent viscosity η_a for a non-Newtonian liquid. The kinematic viscosity is the viscosity coefficient divided by the density, $\nu = \eta / \rho$. The fluidity is the reciprocal of the viscosity, $\phi = 1 / \eta$. The common units for viscosity, dyne seconds per square centimeter ((dyn·s) / cm²) or grams per centimeter second ((g / (cm·s))), called poise, which is usually expressed as centipoise (cP), have been replaced by the SI units of pascal seconds, ie, Pa·s and mPa·s, where 1 mPa·s = 1 cP. In the same manner the shear stress units of dynes per square centimeter, dyn / cm², have been replaced by Pascals, where 10 dyn / cm² = 1 Pa, and newtons per square meter, where 1 N / m² = 1 Pa. Shear rate is $\Delta V / \Delta X$, or length / time / length, so that values are given as per second (s⁻¹) in both systems. The SI units for kinematic viscosity are square centimeters per second, cm² / s, ie, Stokes (St), and square millimeters per second, mm² / s, ie, centistokes (cSt). Information is available for the official Society of Rheology nomenclature and units for a wide range of rheological parameters (11).

Flow Models. Many flow models have been proposed (10,12), which are useful for the treatment of experimental data or for describing flow behavior (Table 1). However, it is likely that no given model fits the rheological behavior of a material over an extended shear rate range. Nevertheless, these models are useful for summarizing rheological data and are frequently encountered in the literature.

Table 1. Flow Equations for Flow Models

Flow model	Flow equation
Newtonian	$\tau = \eta \dot{\gamma}$
plastic (Bingham) body	$\tau - \tau_0 = \eta \dot{\gamma}$
power law	$\tau = k \dot{\gamma} ^n$
power law with yield value	$\tau - \tau_0 = k \dot{\gamma} ^n$
Casson fluid	$\tau^{1/2} = \tau_0^{1/2} + \eta_\infty^{1/2} \dot{\gamma}^{1/2}$
Williamson	$\eta = \eta_\infty + \frac{\eta_0 - \eta_\infty}{1 + \frac{\tau}{\tau_m}}$
cross	$\eta = \eta_\infty + \frac{\eta_0 - \eta_\infty}{1 + a \dot{\gamma}^n}$

Of the models listed in Table 1, the Newtonian is the simplest. It fits water, solvents, and many polymer solutions over a wide strain rate range. The plastic or Bingham body model predicts constant plastic viscosity above a yield stress. This model works for a number of dispersions, including some pigment pastes. Yield stress, τ_0 , and plastic (Bingham) viscosity, $\eta_p = (\tau - \tau_0) / \dot{\gamma}$, may be determined from the intercept and the slope beyond the intercept, respectively, of a shear stress vs shear rate plot.

The other models can be applied to non-Newtonian materials where time-dependent effects are absent. This situation encompasses many technically important materials from polymer solutions to latices, pigment slurries, and polymer melts. At high shear rates most of these materials tend to a Newtonian viscosity limit. At low shear rates they tend either to a yield point or to a low shear Newtonian limiting viscosity. At intermediate shear rates, the power law or the Casson model is a useful approximation.

The power law, $\tau = k \dot{\gamma}^n$, is widely used as a model for non-Newtonian fluids. It holds for many solutions and can describe Newtonian, shear-thinning, and shear-thickening behavior, depending on the power factor, n , also called the flow behavior index. For a Newtonian fluid, $n = 1$ and the equation reduces to the Newtonian model. If n is less than 1, the fluid is shear thinning; if it is greater than 1, the fluid is shear thickening. A test of whether the power law applies and a means of determining n is to plot the log shear stress vs the log shear rate. If the plot is linear, the power law applies. The value of n , which is the reciprocal of the slope of the line, can be used as a measure of the degree of shear thinning or shear thickening. Dividing the power law equation through by $\dot{\gamma}$ gives an expression in terms of viscosity, $\eta = k' \dot{\gamma}^{n-1}$.

The power law model can be extended by including the yield value $\tau - \tau_0 = k \dot{\gamma}^n$, which is called the Herschel-Bulkley model, or by adding the Newtonian limiting viscosity, η_∞ . The latter is done in the Sisko model, $\eta_\infty + k \dot{\gamma}^{n-1}$. These two models, along with the Newtonian, Bingham, and Casson models, are often included in data-fitting software supplied for the newer computer-driven viscometers.

Another model is the Casson equation (13), which is useful in establishing the flow characteristics of inks, paints, and other dispersions. An early form of this expression (eq. 1) was modified (14) to give equation 2.

$$\tau^{1/2} = k_0 + k_1 \dot{\gamma}^{1/2}$$

(1)

$$\eta^{1/2} = \eta_\infty^{1/2} + \tau_0^{1/2} \dot{\gamma}^{-1/2}$$

(2)

The square root of viscosity is plotted against the reciprocal of the square root of shear rate (Fig. 3). The square of the slope is τ_0 , the yield stress; the square of the intercept is η_∞ , the viscosity at infinite shear rate. No material actually experiences an infinite shear rate, but η_∞ is a good representation of the condition where all rheological structure has been broken down. The Casson yield stress τ_0 is somewhat different from the yield stress discussed earlier in that there may or may not be an intercept on the shear stress–shear rate curve for the material. If there is an intercept, then the Casson yield stress is quite close to that value. If there is no intercept, but the material is shear thinning, a Casson plot gives a value for τ_0 that is indicative of the degree of shear thinning.

The Williamson equation is useful for modeling shear-thinning fluids over a wide range of shear rates (15). It makes provision for limiting low and high shear Newtonian viscosity behavior (eq. 3), where τ is the absolute value of the shear stress and τ_m is the shear stress at which the viscosity is the mean of the viscosity limits η_0 and η_∞ , ie, at $\eta = (\eta_0 + \eta_\infty) / 2$.

$$\eta = \eta_\infty + \frac{(\eta_0 - \eta_\infty)}{1 + \frac{\tau}{\tau_m}}$$

(3)

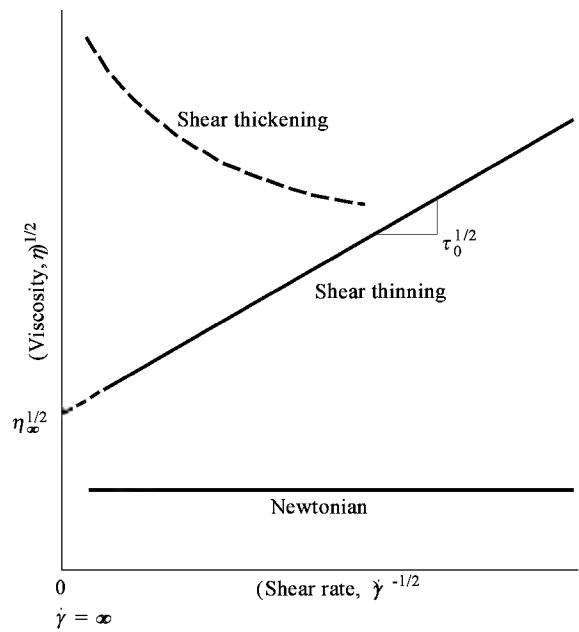


Fig. 3. Examples of Casson plots: (viscosity)^{1/2} vs (shear rate)^{-1/2}. The Casson equation is $\eta^{1/2} = \eta_{\infty}^{1/2} + \tau_0^{1/2} \dot{\gamma}^{-1/2}$.

The Cross equation assumes that a shear-thinning fluid has high and low shear-limiting viscosity (16) (eq. 4), where α and n are constants.

$$\eta = \eta_{\infty} + \left(\frac{\eta_0 - \eta_{\infty}}{1 + \alpha \dot{\gamma}^n} \right) \tag{4}$$

The value for n is often given as 2–3, but polymer melts have shown a wide range of values. The constant α is associated with rupture of the linkages in the structure of the fluid. The effect of different values of α , ie, at the same values of η_0 and η_{∞} , is shown in Figure 4. As α increases, breakdown occurs at lower and lower shear rates.

Thixotropy and Other Time Effects. In addition to the nonideal behavior described, many fluids exhibit time-dependent effects. Some fluids increase in viscosity (rheopexy) or decrease in viscosity (thixotropy) with time when sheared at a constant shear rate. These effects can occur in fluids with or without yield values. Rheopexy is a rare phenomenon, but thixotropic fluids are common. Examples of thixotropic materials are starch pastes, gelatin, mayonnaise, drilling muds, and latex paints. The thixotropic effect is shown in Figure 5, where the curves are for a specimen exposed first to increasing and then to decreasing shear rates. Because of the decrease in viscosity with time as well as shear rate, the up-and-down flow curves do not superimpose. Instead, they form a hysteresis loop, often called a thixotropic loop. Because flow curves for thixotropic or rheoplectic liquids depend on the shear history of the sample, different curves for the same material can be obtained, depending on the experimental procedure.

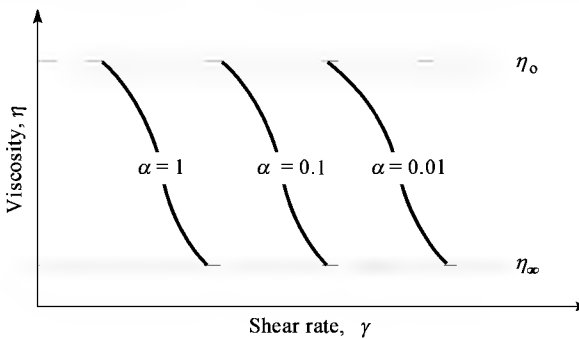


Fig. 4. Graphic representations (viscosity vs shear rate) of Cross model with different values for α .

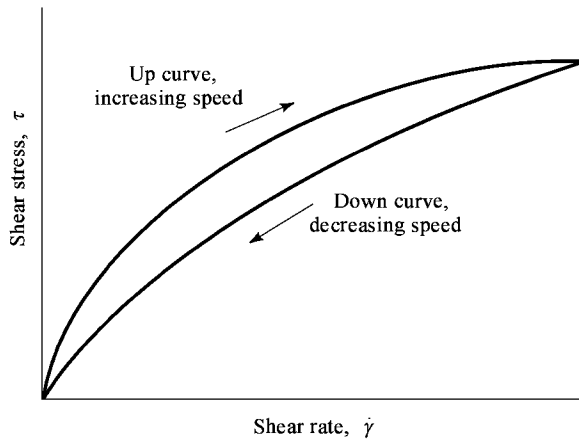


Fig. 5. Flow curves (up and down) for a thixotropic material: hysteresis loop.

Experimentally, it is sometimes difficult to detect differences between a shear-thinning liquid in which the viscosity decreases with increasing shear, and a thixotropic material in which the viscosity decreases with time, because of the combined shear and time effects that occur during a series of measurements. This is especially true if only a few data points are collected. In addition, most materials that are thixotropic are also shear thinning. In fact,

one definition of a thixotropic fluid limits it to materials whose viscosity is a function of both shear rate and time (10). Viscosity–time measurements during or after shearing can be used to show time-dependent effects. The plots in Figure 6 are representative of the results of such measurements on a thixotropic material. The viscosity drops sharply at first, but the rate of change continually decreases until a constant or nearly constant level is reached. This is sometimes referred to as the sheared state. When the shearing action stops, the viscosity increases quickly at first, then increases more slowly, and approaches the original level asymptotically. A good example of this is the behavior of latex house paint. The shearing plot represents the brushing action, whereas the recovery plot shows what happens when the brushing stops. The thixotropic behavior allows the paint to be easily brushed to a thin film and gives a short period of time for the brush marks to level; then the viscosity increase prevents running and sagging.

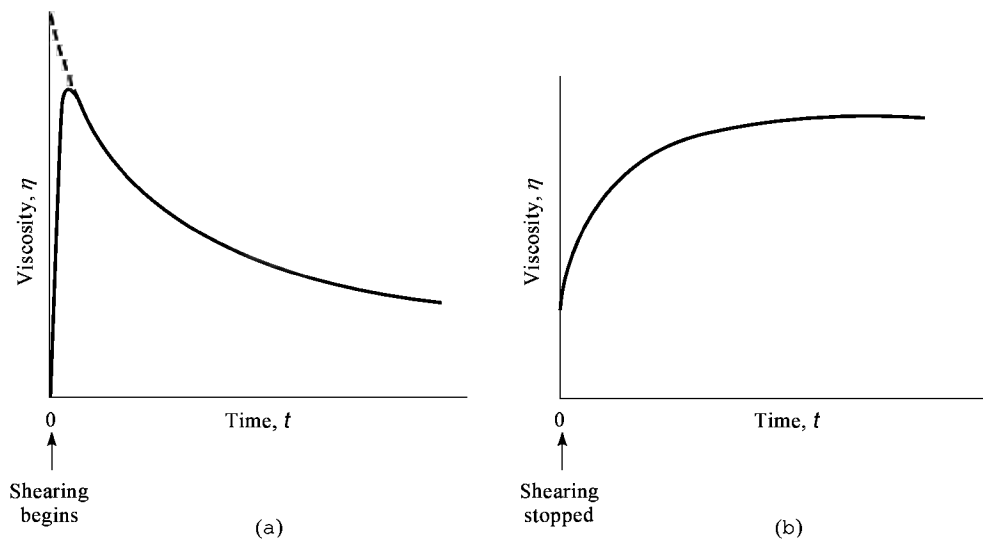


Fig. 6. Viscosity–time effects for a thixotropic material: (a) shearing and (b) recovery. A nonthixotropic material would give horizontal lines in both cases.

Causes of time-dependent behavior include irreversible changes such as cross-linking, coagulation, degradation, and mechanical instability, and reversible changes involving the breaking and reforming of colloidal aggregations and networks. Models of time-dependent behavior are less satisfactory and more controversial than those of shear-dependent behavior. Few comprehensive investigations of the viscosity–shear–time profiles of thixotropic and rheopectic materials have been published, but there are sources that contain good discussions of thixotropy (10,17,18). Rheopexy has also been described (19,20).

Time-dependent effects are measured by determining the decay of shear stress as a function of time at one or more constant shear rates (Fig. 7) (21). A rotational viscometer connected to a recorder is used. After the sample is loaded and allowed to come to mechanical and thermal equilibrium, the viscometer is turned on and the rotational speed is increased in steps, starting from the lowest speed. The resultant shear stress is recorded with time. On each speed change the shear stress reaches a maximum value and then decreases exponentially toward an equilibrium level. The peak shear stress, which is obtained by extrapolating the curve to zero time, and the equilibrium shear stress are indicative of the viscosity–shear behavior of unsheared and sheared material, respectively. The stress–decay curves are indicative of the time-dependent behavior. A rate constant for the relaxation process can be determined at each shear rate. In addition, zero-time and equilibrium shear stress values can be used to construct a hysteresis loop that is similar to that shown in Figure 5, but unlike that plot, is independent of acceleration and time of shear.

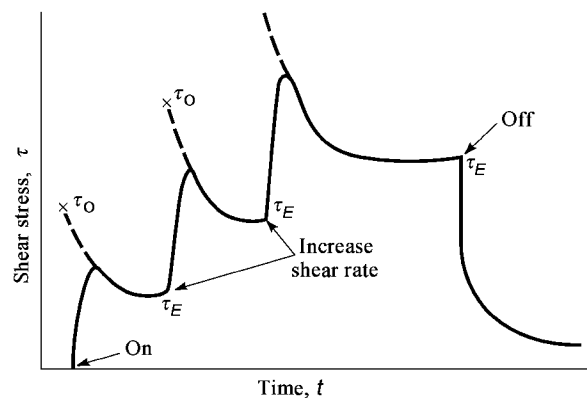


Fig. 7. Decay of shear stress during steady shear at various shear rates. Determination of zero-time shear stresses or yield stresses and equilibrium shear stresses.

Another method is the step-shear test (10), which uses controlled shearing and the recovery behavior shown in Figure 6b to characterize the material. In this method, a high shear rate ($\sim 10^4 \text{ s}^{-1}$) is applied to the specimen until the viscosity falls to an equilibrium value. The shear rate then is reduced to a low value ($\sim 1 \text{ s}^{-1}$), allowing the structure to reform and the viscosity to recover. The data can be analyzed in a number of ways. The time it takes to achieve 50% viscosity recovery or some other fraction of the original value can be used to indicate the rate of recovery. Comparisons can be made based on these times or on the time needed to reach a given viscosity. Equation 5 has been fit to the recovery curve, where $\eta(t)$ is the viscosity as a function of time, $\eta_{t=0}$, the sheared-out viscosity at recovery time zero; $\eta_{t=\infty}$, the infinite time recovered viscosity; and τ , a time constant describing the recovery rate.

$$\eta(t) = \eta_{t=0} + (\eta_{t=\infty} - \eta_{t=0}) \left(1 - e^{-t/\tau}\right) \tag{5}$$

Another method for estimating thixotropy involves the hysteresis of the thixotropic loop. The area of the thixotropic loop is calculated or measured, which works well with printing inks (3). In a variation of this method, the up curve on an undisturbed sample is determined. The sample is then sheared at high shear ($>2000 \text{ s}^{-1}$) for 30–60 s, followed by determination of the down curve (22). The data are plotted as Casson–Asbeck plots, $\eta^{1/2}$ vs $\dot{\gamma}^{1/2}$ (14), as shown in Figure 8. Such plots are best used for comparison and ranking, but a measure of the degree of thixotropy can be gained by determining the angle formed by the two lines or the area of the triangle formed by the lines and a vertical line through a given value for $\dot{\gamma}^{1/2}$. Additional methods for

determining the degree of thixotropy have been described (9).

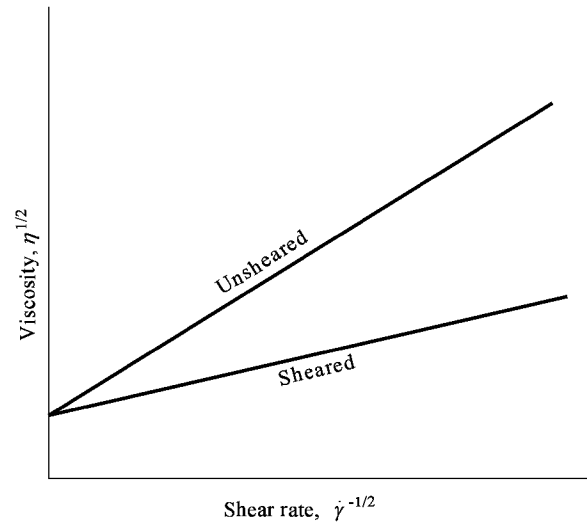


Fig. 8. Casson plots of sheared and unsheared paints. The degree of divergence of the lines is used to estimate thixotropy.

Results from measurements of time-dependent effects depend on the sample history and experimental conditions and should be considered approximate. For example, the state of an unsheared or undisturbed sample is a function of its previous shear history and the length of time since it underwent shear. The area of a thixotropic loop depends on the shear range covered, the rate of shear acceleration, and the length of time at the highest shear rate. However, measurements of time-dependent behavior can be useful in evaluating and comparing a number of industrial products and in solving flow problems.

Effect of Temperature. In addition to being often dependent on parameters such as shear stress, shear rate, and time, viscosity is highly sensitive to changes in temperature. Most materials decrease in viscosity as temperature increases. The dependence is logarithmic and can be substantial, up to 10° change °C. This has important implications for processing and handling of materials and for viscosity measurement.

A common expression relating viscosity to temperature is the Arrhenius equation, $\eta = A \cdot e^{B/T}$ or $\eta = A \cdot 10^{B/T}$, where *A* and *B* are constants characteristic of the polymer or other material, and *T* is the absolute temperature. Estimation of the viscosity of a polymer at a given temperature requires a knowledge of the viscosity at two other temperatures. This knowledge allows calculation of the constants *A* and *B* and subsequent determination of viscosities at other, intermediate temperatures.

The Arrhenius equation may also be expressed in logarithmic form (eq. 6):

$$\log \eta = \log A + B/T$$

(6)

For two temperatures (viscosity known at one), this expression may be written as (eq. 7):

$$\log(\eta_1/\eta_2) = B \left(\frac{1}{T_1} - \frac{1}{T_2} \right)$$

(7)

In practice, plots of log η vs $1/T$ tend to be straight lines over a considerable range of values. This permits the construction of nomographs (23). The viscosity at an intermediate temperature is determined from the known viscosities and two other temperatures by drawing straight lines connecting the viscosities and corresponding temperatures on the respective scales. The intersection of the two straight lines gives the pivot point. The viscosity at an intermediate temperature can be determined by drawing a line through the temperature and pivot point to the viscosity scale.

The Arrhenius equation holds for many solutions and for polymer melts well above their glass-transition temperatures. For polymers closer to their *T_g* and for concentrated polymer and oligomer solutions, the Williams-Landel-Ferry (WLF) equation (24) works better (25,26). With a proper choice of reference temperature *T_r*, the ratio of the viscosity to the viscosity at the reference temperature can be expressed as a single universal equation (eq. 8):

$$\log(\eta/\eta_s) = \frac{8.86(T - T_s)}{101.6 + (T - T_s)}$$

(8)

Because *T_r* is often defined as *T_g* + 50 K, the equation thus becomes

$$\log(\eta/\eta_{T_g}) = \frac{17.4(T - T_g)}{51.6 + (T - T_g)}$$

In general, the WLF equation holds over the temperature range *T_g* to (*T_g* + 100°C).

Dilute Polymer Solutions. The measurement of dilute solution viscosities of polymers is widely used for polymer characterization. Very low concentrations reduce intermolecular interactions and allow measurement of polymer–solvent interactions. These measurements are usually made in capillary viscometers, some of which have provisions for direct dilution of the polymer solution. The key viscosity parameter for polymer characterization is the limiting viscosity number or intrinsic viscosity, $[\eta]$. It is calculated by extrapolation of the viscosity number (reduced viscosity) or the logarithmic viscosity number (inherent viscosity) to zero concentration.

The viscosity ratio or relative viscosity, η_{rel} , is the ratio of the viscosity of the polymer solution to the viscosity of the pure solvent. In capillary viscometer measurements, the relative viscosity (dimensionless) is the ratio of the flow time for the solution *t* to the flow time for the solvent *t₀* (Table 2). The specific (sp) viscosity (dimensionless) is also defined in Table 2, as is the viscosity number or reduced (red) viscosity, which has the units of cubic meters per kilogram (m³/kg) or deciliters per gram (dL/g). The logarithmic viscosity number or inherent (inh) viscosity likewise has the units m³/kg or dL/g. For η_{red} and η_{inh} , *c*, the concentration of polymer, is expressed in convenient units, traditionally g/100 cm³ but kg/m³ in SI units. The viscosity number and logarithmic viscosity number vary with concentration, but each can be extrapolated (Fig. 9) to zero concentration to give the limiting viscosity number (intrinsic viscosity) (Table 2).

Table 2. Viscosity Expressions

Common name	Recommended name	Definition	Common
-------------	------------------	------------	--------

			units
relative viscosity	viscosity ratio	$\eta_{rel} = t/t_0 = \eta/\eta_0$	none
specific viscosity		$\eta_{sp} = (\eta - \eta_0)/\eta_0 = \eta_{rel} - 1$	none
reduced viscosity	viscosity number	$\eta_{red} = \eta_{sp}/c = (\eta_{rel} - 1)/c$	dL/g
inherent viscosity	logarithmic viscosity number	$\eta_{inh} = (\ln \eta_{rel})/c$	dL/g
intrinsic viscosity	limiting viscosity number	$[\eta] = \lim_{c \rightarrow 0} \frac{\eta_{sp}}{c} = \lim_{c \rightarrow 0} \frac{\ln \eta_{rel}}{c}$	dL/g

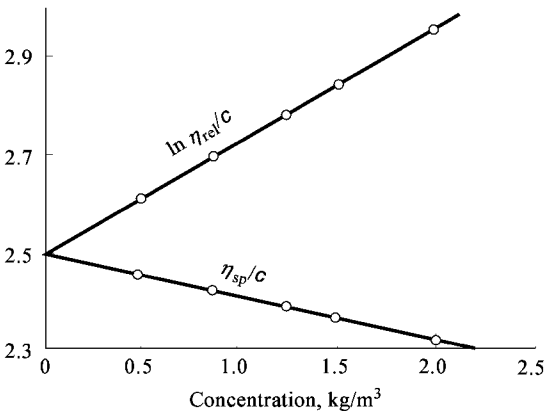


Fig. 9. Plots of viscosity number ($\eta_{red} = \eta_{sp}/c$) and the logarithmic viscosity number ($\eta_{inh} = \ln \eta_{rel}/c$) vs concentration. Extrapolations to zero concentration give the limiting viscosity number $[\eta]$.

Extrapolation to infinite dilution requires viscosity measurements at usually four or five concentrations. For relative (rel) measurements of rapid determination, a single-point equation may often be used. A useful expression is the following (eq. 9) (27):

$$[\eta] = 2(\eta_s - \ln \eta_{rel})^{-1/2} c$$

(9)

The general validity of single-point methods has been questioned, however (28). An even simpler but equally useful method is to approximate the limiting viscosity number by the logarithmic viscosity number of a single sufficiently dilute solution: $[\eta] \sim \eta_{inh} = (\ln \eta_{rel})/c$, where $c = 1\text{--}2 \text{ kg m}^{-3}$ or $0.1\text{--}0.2 \text{ g } 100 \text{ cm}^{-3}$.

The limiting viscosity number or intrinsic viscosity is an important measure of the characteristics of the polymer and of its interactions with the solvent in which it is dissolved. The limiting viscosity number depends on the polymer, solvent, and temperature, but under a given set of conditions it is related to the molecular weight by the Mark-Houwink relation, $[\eta] = KM^a$, where K and a are constants and M is the molecular weight of the polymer. Tables of K and a are available for a large number of polymers and solvents (29,30). Excellent summaries of equations, techniques, and references relating to the viscosity of dilute polymer solutions are also available (31,32), as is information on dilute polymer solutions that are shear thinning (33).

Concentrated Polymer Solutions. Knowledge of the viscosity behavior of concentrated solutions (34–36) is important to the manufacture and application of caulks, adhesives, inks, paints, and varnishes. It is also useful for designing and controlling polymer manufacturing processes, fiber spinning, and film casting. Viscosity behavior can be investigated by a variety of methods, including the use of simple capillary viscometers, extrusion rheometers, and rotational viscometers. Unlike dilute solutions, concentrated polymer solutions show a vast amount of interaction between the macromolecules. The degree of interaction is governed by the concentration, the characteristics of the chains, and the nature of the solvent. A convenient measure of concentration is the dimensionless reduced concentration c^* , which is the product of the concentration and the limiting viscosity number (intrinsic viscosity) $[\eta]$ (34). The transition from dilute to concentrated solutions occurs at a critical concentration c_c and corresponds to c^* values of several units. In addition, at a critical molecular weight, where $M > M_c$ and $c > c_c$, a fluctuating entanglement network forms. For a concentrated solution, properties above and below M_c may be quite different. For example, the dependence of viscosity on molecular weight, which is much greater in concentrated than in dilute solutions, changes from a value on the order of unity below M_c to one of 3.4–3.5 above M_c (35,37). That is, $\eta = KM$ below M_c and $\eta = KM^{3.4\text{--}3.5}$ above M_c . Viscosity in these expressions should be the zero-shear viscosity η_0 , but because the relationships hold for low shear measurements in many cases, the notation remains in the more general form η .

The break point above which entanglement occurs varies widely with molecular structure; a range of 3,800–36,000 mol wt has been shown (38). In highly concentrated oligomeric solutions such as high solids organic coatings (volume fraction >0.7), high dependencies of viscosity on molecular weight occur even at low molecular weights (39,40). This is probably on account of hydrogen bonding that causes the stringing together of short chains or the formation of a loose network, thereby increasing the effective chain length. Something similar to this has been seen with the formation of viscosity-building needle-like structures by low molecular weight materials (41).

Depending on the concentration, the solvent, and the shear rate of measurement, concentrated polymer solutions may give wide ranges of viscosity and appear to be Newtonian or non-Newtonian. This is illustrated in Figure 10, where solutions of a styrene–butadiene–styrene block copolymer are Newtonian and viscous at low shear rates, but become shear thinning at high shear rates, dropping to relatively low viscosities beyond 10^5 s^{-1} (42). The shear rate at which the break in behavior occurs depends on the concentration and on the solvent.

Melt Viscosity. The study of the viscosity of polymer melts (43–55) is important for the manufacturer who must supply suitable materials and for the fabrication engineer who must select polymers and fabrication methods. Thus melt viscosity as a function of temperature, pressure, rate of flow, and polymer molecular weight and structure is of considerable practical importance. Polymer melts exhibit elastic as well as viscous properties. This is evident in the swell of the polymer melt upon emergence from an extrusion die, a behavior that results from the recovery of stored elastic energy plus normal stress effects.

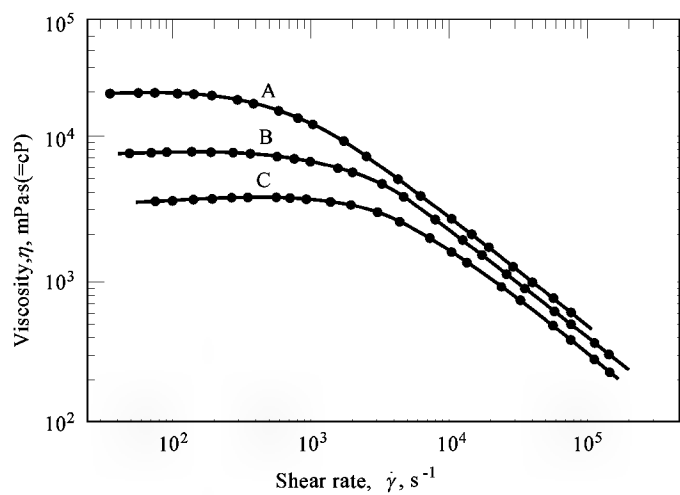


Fig. 10. Viscosity vs shear rate for solutions of a styrene–butadiene–styrene block copolymer (42). A represents cyclohexanone, where $c = 0.248 \text{ g cm}^{-3}$; B, o -xylene, where $c = 0.246 \text{ g cm}^{-3}$; C, toluene, where $c = 0.248 \text{ g cm}^{-3}$. Courtesy of the Society of Plastics Engineers, Inc.

A number of experimental methods have been applied to measure the melt viscosity of polymers (49,51), but capillary extrusion techniques probably are generally preferred. Rotational methods are also used, and some permit the measurement of normal stress effects resulting from elasticity as well as of viscosity. Slit rheometers can also be used to measure normal stress. Oscillatory shear measurements are useful for measuring the elasticity of polymer melts (52,53). Controlled stress methods have also been applied (54). Squeeze film flow has also been proposed as a geometry suitable for processability testing of polymer melts (55). Nonlinear viscoelastic behavior is found in many molten plastics. This has been addressed by various technologies, including a sliding plate normal-thrust rheometer (56).

Polymer melts show a low shear rate Newtonian limit and a region of diminishing viscosity with increasing shear rate. Although it is likely that a high shear rate Newtonian region exists, this has generally not been observed experimentally because of the effects of heat generation and polymer degradation at high shear rates.

The limiting low shear or zero-shear viscosity η_0 of the molten polymer can be related to its weight-average molecular weight, $\overline{M}_w$, by the same relations noted for concentrated solutions: $\eta_0 = K \overline{M}_w$ for low molecular weight and $\eta_0 = K \overline{M}_w^{1.5}$ for high molecular weight.

The transition between two forms of behavior occurs at a critical molecular weight M_c , which corresponds to a critical chain length, Z_c . The transition is clearly shown in Figure 11, which is a plot of Newtonian viscosity vs chain length in terms of carbon atoms for a series of molten polyethylenes (34). This transition is thought to be related to chain fluctuations in the polymer melt in such a way that the motion changes from displacement of whole chains to restricted motion called reptation, ie, wiggly, snake-like motion within the tight tube formed by the matrix of neighboring chains (57–60). Polymer entanglement and reptation have considerable influences on melt rheology, particularly on viscoelasticity, with many consequences for the molding and extrusion of plastics.

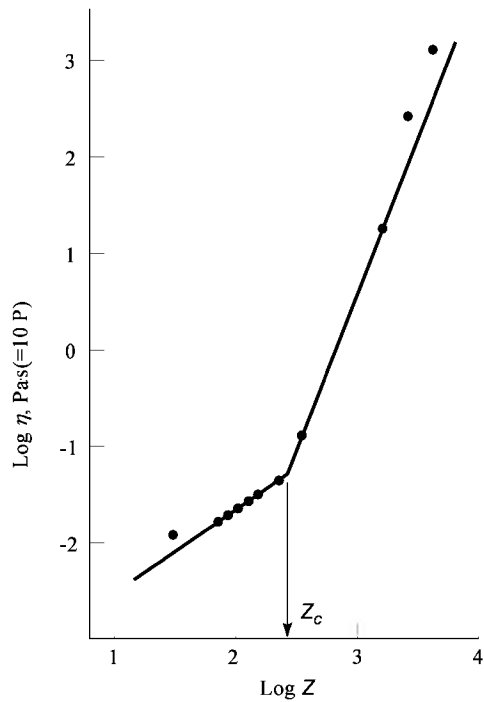


Fig. 11. Newtonian viscosity vs chain length in terms of the number of carbon atoms for a series of molten polyethylenes. To convert Pa·s to P, multiply by 10.

Courtesy of Springer-Verlag.

The viscosity–molecular weight relationships noted above hold for narrow molecular weight distribution polymers. For polymers having broad molecular weight distributions, viscosity depends on a molecular weight average between $\overline{M}_w$ and the next higher average $\overline{M}_z$ ($\overline{M}_z$ average), approaching $\overline{M}_z$ as the distribution broadens. Branching can have a considerable effect, reducing or increasing the viscosity depending on the ratio of the length of the branch to the entanglement length (61). The flow curve for a polymer melt can be predicted from the molecular weight distribution (62) and, under some circumstances, a molecular weight distribution can be determined from the flow curve (63,64). As might be expected, fillers also have an effect on melt viscosity (65).

The dependence of viscosity on temperature is critical to the handling of molten polymers in molding, extrusion, and other manufacturing processes. In fact, the drop in viscosity with increasing temperature makes these operations possible. Therefore, viscosity–temperature relationships are

important. Data for many polymers can be found in the literature (43–51). The behavior of others can be determined by viscosity measurements over a range of temperatures. Viscosity–stress master curves are useful in the prediction of viscosity at a given temperature (66). The viscosity–stress curves shown in Figure 12 (43) are approximately superimposable by shifting at constant stress.

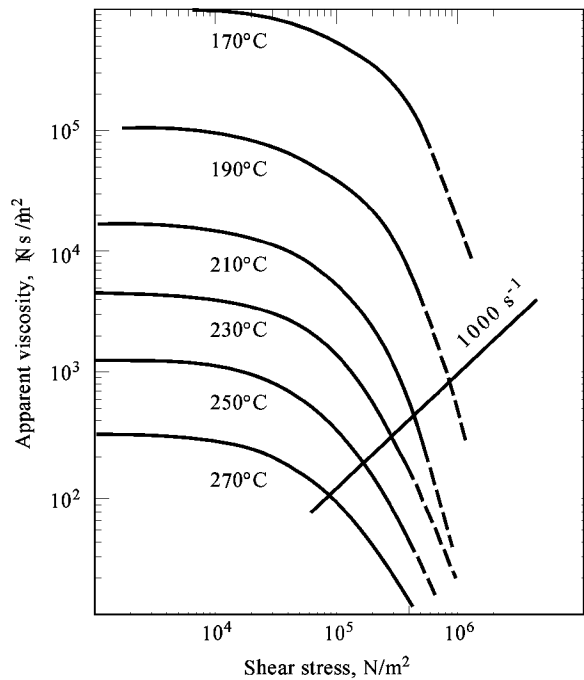


Fig. 12. Viscosity at different temperatures measured by a capillary viscometer: injection-molding grade of poly(methyl methacrylate) (43). To convert N m^{-2} to psi, multiply by 145; to convert $(\text{N}\cdot\text{s}) \text{ m}^{-2}$ to $(\text{dyn}\cdot\text{s}) \text{ cm}^{-2}$ (P), multiply by 10.

The temperature dependence of melt viscosity at temperatures considerably above T_g approximates an exponential function of the Arrhenius type. However, near the glass transition the viscosity temperature relationship for many polymers is in better agreement with the WLF treatment (24).

Melt viscosity is also affected by pressure (43,67,68). The compression of a melt reduces the free volume and therefore raises the viscosity. For example, the viscosity of low density polyethylene increases by a factor of roughly 10 over a static pressure range of 34–170 MPa (5,000–25,000 psi).

Dispersed Systems. Many fluids of commercial and biological importance are dispersed systems, such as solids suspended in liquids (dispersions) and liquid–liquid suspensions (emulsions). Examples of the former include inks, paints, pigment slurries, and concrete; examples of the latter include mayonnaise, butter, margarine, oil-and-vinegar salad dressing, and milk. Blood seems to fall in between as it is a suspension of deformable but not liquid particles, and it does not behave like either a dispersion or an emulsion (69); it thus has an interesting rheology (70).

Dispersion of a solid or liquid in a liquid affects the viscosity. In many cases Newtonian flow behavior is transformed into non-Newtonian flow behavior. Shear thinning results from the ability of the solid particles or liquid droplets to come together to form network structures when at rest or under low shear. With increasing shear the interlinked structure gradually breaks down, and the resistance to flow decreases. The viscosity of a dispersed system depends on hydrodynamic interactions between particles or droplets and the liquid, particle–particle interactions (bumping), and interparticle attractions that promote the formation of aggregates, flocs, and networks.

Emulsions have not been studied as thoroughly as dispersions, probably because of the greater complexity of the former. Emulsions tend to be unstable, and frequently droplets begin to coalesce soon after the emulsion forms. Thus the emulsion is continually changing. In addition, droplets may change shape under shear or when packed tightly together. This makes it difficult to apply theories developed for solid spheres or other well-defined geometries of shape.

Detailed treatments of the rheology of various dispersed systems are available (71–73), as are reviews of the viscous and elastic behavior of dispersions (74,75), of the flow properties of concentrated suspensions (75–82), and of viscoelastic properties (83–85). References are also available that deal with blood red cell suspensions (69,70,86).

Viscosity–Concentration Relationship for Dilute Dispersions. The viscosities of dilute dispersions have received considerable theoretical and experimental treatment, partly because of the similarity between polymer solutions and small particle dispersions at low concentration. Nondeformable spherical particles are usually assumed in the cases of molecules and particles. The key viscosity quantity for dispersions is the relative viscosity or viscosity ratio, η_{rel} .

$$\eta_{rel} = \frac{\eta}{\eta_0} = \frac{\text{dispersion viscosity}}{\text{viscosity of liquid}}$$

This is because the effect of the dispersed solid, rather than the dispersing medium, is usually more significant. However, the latter should not be ignored. Many industrial problems involving unacceptably high viscosities in dispersed systems are solved by substituting solvents of lower viscosity.

The relative viscosity of a dilute dispersion of rigid spherical particles is given by $\eta_{rel} = 1 + a\phi$, where a is equal to $[\eta]$, the limiting viscosity number (intrinsic viscosity) in terms of volume concentration, and ϕ is the volume fraction. Einstein has shown that, provided that the particle concentration is low enough and certain other conditions are met, $[\eta] = 2.5$, and the viscosity equation is then $\eta_{rel} = 1 + 2.5\phi$. This expression is usually called the Einstein equation.

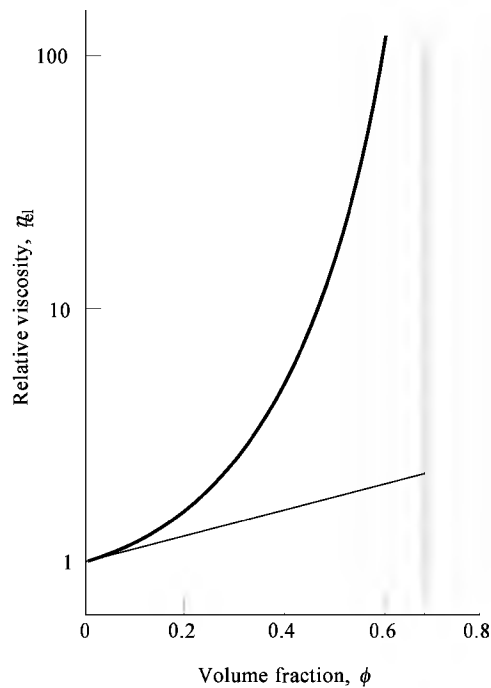


Fig. 13. Relative viscosity vs volume fraction for a typical dispersion (—), where (---) represents the Einstein relationship $[\eta] = 2.5$, and (---) is an approximate value for the limiting volume fraction ϕ_m .

For higher ($\phi > 0.05$) concentrations where particle–particle interactions are noticeable, the viscosity is higher than predicted by the Einstein equation. The viscosity–concentration equation becomes equation 10, where b and c are additional constants (87).

$$\eta_{rel} = 1 + 2.5\phi + b\phi^2 + c\phi^3 + \dots \tag{10}$$

The deviation from the Einstein equation at higher concentrations is represented in Figure 13, which is typical of many systems (88,89). The relative viscosity tends to infinity as the concentration approaches the limiting volume fraction of close packing ϕ_m ($\phi \sim 0.7$). Equation 10 has been modified (90,91) to take this into account, and the expression for η_{rel} becomes (eq. 11):

$$\eta_{rel} = \frac{1 + 2.5\phi + b\phi^2 + c\phi^3 + \dots}{1 - (\phi_m - \phi)} \tag{11}$$

Another model that has been successful in fitting much data is the Krieger-Dougherty equation (eq. 12) (10,92):

$$\eta_{rel} = \left(1 - \frac{\phi}{\phi_m}\right)^{[\eta]\phi_m} \tag{12}$$

Other Factors Affecting the Viscosity of Dispersions. Factors other than concentration affect the viscosity of dispersions. A dispersion of nonspherical particles tends to be more viscous than predicted if the Brownian motion is great enough to maintain a random orientation of the particles. However, at low temperatures or high solvent viscosities, the Brownian motion is small and the particle alignment in flow (streamlining) results in unexpectedly lower viscosities. This is a form of shear thinning.

If the dispersion particles are attracted to each other, they tend to flocculate and form a structure. At low concentrations the particles form open aggregates, which give a fractal structure (93,94). At higher concentrations a network structure results, which can be so pronounced that the mixture has a yield point and behaves like a solid when at rest. Shearing breaks up this structure, and viscosity decreases.

If the volume concentration of the solid in the dispersion is high enough, shearing may produce an increase rather than a decrease in viscosity. Such behavior, called shear thickening (10,17,95,96) or dilatancy, is common in dispersions of certain pigments and other powders. These dispersions are closely packed, but are usually shear thinning up to moderate shear rates, ie, a few hundred to a few thousand s^{-1} . Higher shear introduces irregularities in the packing, with bridging effects occurring between the particles. The overall packing loosens, which implies that the total space between particles increases. The liquid is not able to fill this space and can no longer wet all the particles. Much of the lubricating effect of the liquid is therefore lost, and internal friction rises, producing a high viscosity. This is true dilatancy, and the volume actually increases. Shear thickening can also occur in dilute suspensions (97). In such cases there is no volume expansion; instead there is buildup of particle aggregates which ultimately produces a network that behaves like a gel.

Emulsions. Because emulsions are different from dispersions, different viscosity–concentration relationships must be used (71,87). In an emulsion the droplets are not rigid, and viscosity can vary over a wide range. Several equations have been proposed to account for this. An extension of the Einstein equation includes a factor that allows for the effect of variations in fluid circulation within the droplets and subsequent distortion of flow patterns (98,99).

Extensional Viscosity. In addition to the shear viscosity η , two other rheological constants can be defined for fluids: the bulk viscosity, K , and the extensional or elongational viscosity, η_e (34,49,100–107). The bulk viscosity relates the hydrostatic pressure to the rate of deformation of volume, whereas the extensional viscosity relates the tensile stress to the rate of extensional deformation of the fluid. Extensional viscosity is important in a number of industrial processes and problems (34,100,108–110). Shear properties alone are insufficient for the characterization of many fluids, particularly polymer melts (101,107,111,112).

Extensional flows occur when fluid deformation is the result of a stretching motion. Extensional viscosity is related to the stress required for the stretching. This stress is necessary to increase the normalized distance between two material entities in the same plane when the separation is s and the relative velocity is ds/dt . The deformation rate is the extensional strain rate, which is given by equation 13 (108):

$$\dot{\epsilon} = \frac{1}{s} \frac{ds}{dt} \tag{13}$$

Unlike shear viscosity, extensional viscosity has no meaning unless the type of deformation is specified. The three types of extensional viscosity identified and measured are uniaxial or simple, biaxial, and pure shear. Uniaxial viscosity is the only one used to characterize fluids. It has been employed mainly in the study of polymer melts, but also for other fluids. For a Newtonian fluid, the uniaxial extensional viscosity is three times the shear viscosity:

$(\eta_e)_{lim} = 3\eta$. This is confirmed at very low shear rates in Figure 14, which provides a typical example of the extensional viscosity behavior of a polymer (111). The two other extensional viscosities are used to study elastomers in the form of films or sheets. Uniaxial and biaxial extensions are important in industry (100,107–110,113,114), the former for the spinning of textile fibers and roller spattering of paints, and the latter for blow molding, vacuum forming, film blowing, and foam processes.

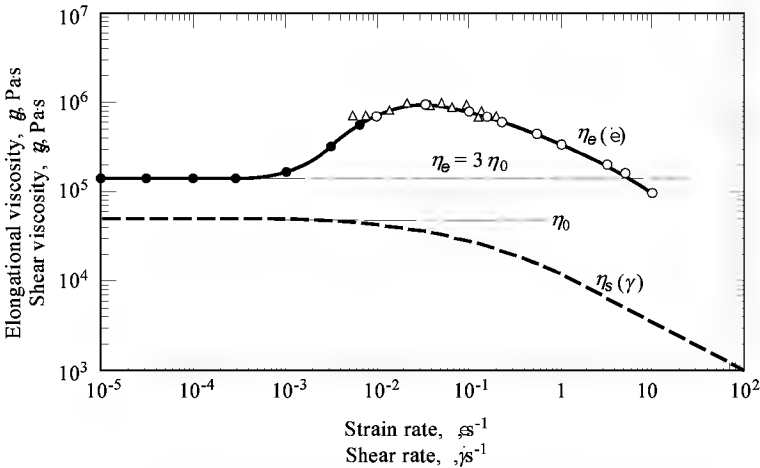


Fig. 14. Shear viscosity, η_s , and extensional viscosity, η_e , as a function of deformation rate of a low density polyethylene (LDPE) at 150°C (111). To convert Pa·s to P, multiply by 10.

Courtesy of *Rheologica Acta*.

Electrorheological Behavior. Electrorheological (ER) fluids are colloidal suspensions whose properties change strongly and reversibly upon application of an electric field. The first detailed study was published in 1949 (115), but ER fluids remained more of a curiosity than useful materials until the 1990s. Interest in ER fluids has increased greatly because they appear to offer a means for rapid processes or operations that take advantage of the calculation and control speeds of modern computers. Manufacturers hope to use these smart materials in applications such as fast acting switches and valves, robot operations, safety catches and other mechanical release devices, vibration damping, automobile suspensions that change with road conditions, variable speed drives, and clutches. Design and fabrication of such devices requires an understanding of the nature and flow properties of ER materials.

When an electric field is applied to an ER fluid, it responds by forming fibrous or chain structures parallel to the applied field. These structures greatly increase the viscosity of the fluid, by a factor of 10^5 in some cases. At low shear stress the material behaves like a solid. The material has a yield stress, above which it will flow, but with a high viscosity. The force necessary to shear the fluid is proportional to the square of the electric field (116).

A parameter used to characterize ER fluids is the Mason number, M_a , which describes the ratio of viscous to electrical forces, and is given by equation 14, where ϵ is the solvent dielectric constant; η_0 , the solvent viscosity; $\dot{\gamma}$, the strain or shear rate; β , the effective polarizability of the particles; and E , the electric field (117).

$$M_a = \frac{24\pi\epsilon\eta_0\dot{\gamma}}{(\beta E)^2}$$

(14)

Many investigators believe that the Bingham model accounts best for observations of electrorheological behavior (116,118), but other models have also been proposed (116,119). There is considerable evidence that ER materials behave as linear viscoelastic fluids while under the influence of electric field (120); thus it appears that these materials may be thought of as elastic Bingham fluids.

The literature on ER fluids and their rheology has been expanding rapidly; a selection of reviews and papers is given in References 116–139. An entire issue of the *Journal of Rheology* has been committed to electrorheology (140), which provides a good survey of the subject at that time. There is an analogous magnetorheological (MR) effect where suspensions filled with magnetic particles show reversible changes in their rheological behavior when subjected to a magnetic field (141,142).

Elasticity and Viscoelasticity

Elastic deformation is a function of stress and is expressed in terms of relative displacement or strain. Strain may be expressed in terms of relative change in volume, length, or other measurement, depending on the nature of the stress. An ideal elastic body (143,144) is a material for which the strain is proportional to the stress (Hooke's law) with immediate recovery to the original volume and shape when the stress is released. The relationship between the stress σ and strain ϵ may be written as $\sigma = K\epsilon$, where K is a proportionality constant called the modulus of elasticity. For a homogeneous, isotropic, Hookean solid, three moduli may be defined. Young's modulus, E , relates tensile stress to tensile strain. The shear modulus, G , relates shear stress to shear strain, ie, $G = \tau / \gamma$. The bulk modulus B relates hydrostatic pressure to the change in volume. Another elastic constant needed for complete specification of behavior in tension is Poisson's ratio, μ , which relates change in volume to change in shape. Incompressible solids and polymer melts have $\mu = 0.5$, but for most solid materials, $\mu < 0.5$. For isotropic, Hookean materials, Young's modulus is related to the shear modulus by $E = 2G(1 + \mu)$. If μ is 0.5, Young's modulus is three times the shear modulus. Values for the various moduli and Poisson's ratio for some representative materials are given in Table 3.

Materials such as metals are nearly elastic and show almost no flow or viscous component. Polymers and many of their solutions are both viscous and elastic, and both types of deformation must be taken into account to explain their behavior.

Mechanical Behavior of Materials. Different kinds of materials respond differently when they undergo basic mechanical tests. This is illustrated in Figure 15, which shows stress–strain diagrams for purely viscous and purely elastic materials. With the former, the stress is relieved by viscous flow and is independent of strain. With the latter, there is a direct dependence of stress on strain and the ratio of the two is the modulus E (or G).

Table 3. Measured Values of Elastic Constants at Small Extensions and 25°C^a

Material	Young's modulus,	Proportionality limit ^c , % extension	Shear modulus, G , GPa ^b	Poisson's ratio, μ	Bulk modulus, B , GPa ^b
	E , GPa ^b				
vitreous silica	70.0	3	30.5	0.14	37.4
mild steel	200–220	2.5	76–83	0.29	160.0
brass	80–100	2	26–38	0.25–0.4	61
constantan	163	2	61	0.33	160.0
nickel	200–220	2	78–80	0.30	170.0

tin	39–55		17–20	0.33	52
silver	60–80	2	24–28	0.38	100
granite	ca 30	0.5	ca 10	ca 0.3	ca 30
gelatin gel ^d	2×10^{-4}	10		0.50	
dry wood					
axial piece	4–18	1			
radial piece	ca 1	1			
silk thread	6.4	1			
natural rubber	8.6×10^{-4}	400–600	2.9×10^{-4}	0.49	0.019
hard rubber	0.36	3			
phenolic resin, mineral-filled	2.4	2			
nylon	0.31	3			
polyethylene	0.010	2			

^a Ref. 21.
^b To convert GPa to psi, multiply by 145,000.
^c Values are approximate.
^d 80° H₂O.

The response of different materials to a stress applied at time $t = t_0$ followed by removal of that stress at $t = t_1$, ie, creep and recovery, is shown in Figure 16 (145). For an elastic material (Fig. 16a), the resulting strain is instantaneous and constant until the stress is removed, at which time the material recovers and the strain immediately drops back to zero. In the case of the viscous fluid (Fig. 16b), the strain increases linearly with time. When the load is removed, the strain does not recover but remains constant. Deformation is permanent. The response of the viscoelastic material (Fig. 16c) draws from both kinds of behavior. An initial instantaneous (elastic) strain is followed by a time-dependent strain. When the stress is removed, the initial strain recovery is elastic, but full recovery is delayed to longer times by the viscous component.

Mechanical Models. Because the complex rheological behavior of viscoelastic bodies is difficult to visualize, mechanical models are often used. In these models the viscous response to applied stress is assumed to be that of a Newtonian fluid and is represented by a dashpot, ie, a piston operating in a cylinder of Newtonian fluid. The elastic response is idealized as an ideal elastic (Hookean) solid and is represented by a spring. The dashpot represents the dissipation of energy in the form of heat, whereas the spring represents a system that stores energy. Mechanical behavior of materials may be approximated by combinations of springs and dashpots. These models, which are largely being replaced by mathematical (integral and differential) models, have been covered in detail elsewhere (10,12,21,87,145,146).

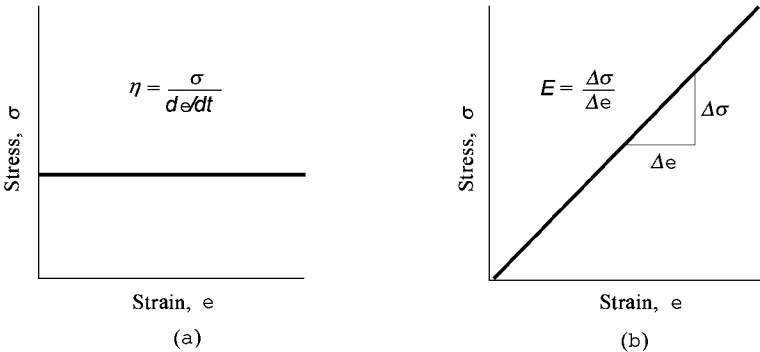


Fig. 15. Stress–strain diagrams. (a) Viscous material of viscosity η ; the stress is independent of strain, but dependent on the speed of testing. (b) Elastic material of modulus E ; the slope is the modulus which is independent of the speed of testing.

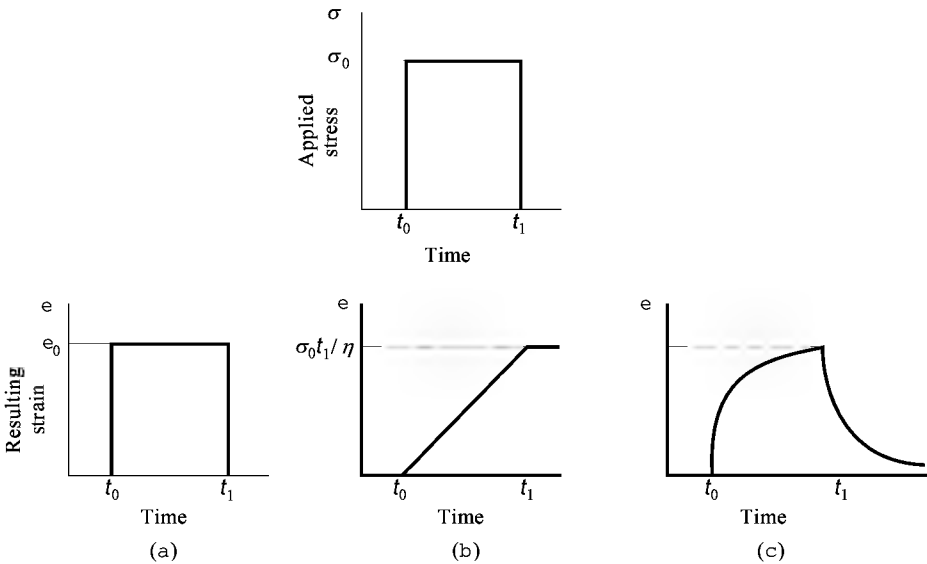


Fig. 16. Response (strain) of different idealized materials to an instantaneous application of a stress at time $t = t_0$: (a) elastic, (b) viscous, and (c) viscoelastic.

Whether a viscoelastic material behaves as a viscous liquid or an elastic solid depends on the relation between the time scale of the experiment and the time required for the system to respond to stress or deformation. Although the concept of a single relaxation time is generally inapplicable to real materials, a mean characteristic time can be defined as the time required for a stress to decay to 1/e of its elastic response to a step change in strain. The

ratio of this characteristic time to the time scale of the experiment, t_{δ} is called the Deborah number. A material at a high Deborah number responds elastically, whereas that at a low Deborah number exhibits viscous behavior: at $t_e \ll \lambda$, it behaves like an elastic solid; at $0t_e > \lambda$, like a viscous liquid. These effects can be seen in geological strata, where rock has flowed to relieve the stresses imposed by geological events. The time scale is long, and the material appears to be viscous.

Dynamic Behavior. Knowledge of the response of materials to stress– strain, creep, and stress–relaxation measurements are useful to define material properties. The dynamic response of viscoelastic materials to cyclic stresses or strains is also important, partly because cyclic motion occurs in many processing operations and applications, and partly because so much rheological information can be gained from dynamic measurements. By subjecting a specimen to oscillatory stress and determining the response, both the elastic and viscous or damping characteristics can be obtained. Elastic materials store energy; that is, they convert mechanical work into potential energy, which is recoverable. Liquids do not store energy when stressed, but dissipate it as heat when they flow. This dissipation results in highly damped motion. Viscoelastic materials exhibit both elastic and damping behavior. The latter causes the deformation to be out of phase with the stress applied in the dynamic measurement.

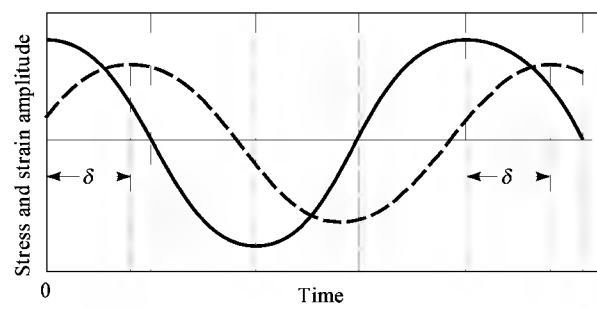


Fig. 17. Viscoelastic material: stress (—) and strain (- - -) amplitudes vs time where δ is the phase angle that defines the lag of the strain behind the stress.

A sinusoidal stress applied to an ideal elastic material produces a sinusoidal strain proportional to the stress amplitude and in phase with it. For ideal viscous materials the stress and strain are out of phase by 90° . Figure 17 gives an example of a stress–strain diagram for a sinusoidal stress applied to a real material. The amplitude of the deformation (strain) in response to the stress is proportional to that of the stress, but lags behind the strain curve by some angle δ between 0 and 90° , depending on whether the material is elastic, viscous, or both. A Hookean (ideal elastic) solid gives an immediate deformation that snaps back when the stress reverses. The spring is completely and instantaneously reversible. The stress and strain are in phase with each other, and the phase angle $\delta = 0$. With a viscous or viscoelastic material, deformation is delayed and then continues in the original direction for a time after the stress reverses. The net result is that the deformation is always out of phase. The less elastic and more ideally viscous the fluid, the greater the lag. A Newtonian fluid such as water gives the greatest lag with a phase angle of 90° .

This behavior is usually analyzed by setting up what are known as complex variables to represent stress and strain. These variables, complex stress and complex strain, ie, τ^* and γ^* , respectively, are vectors in complex planes. They can be resolved into real (in phase) and imaginary (90° out of phase) components similar to those for complex modulus shown in Figure 18.

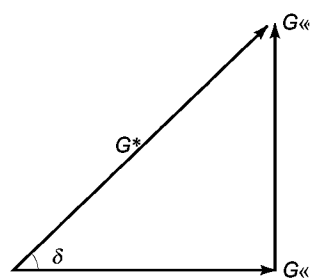


Fig. 18. Resolution of the complex modulus G^* into two vectors, G' , the storage modulus, and G'' , the loss modulus; the phase angle is δ .

The complex stress is $\tau^* = \tau' + i\tau''$, which is the sum of a real part of the stress and an imaginary part; the complex strain is $\gamma^* = \gamma' + i\gamma''$, where i is the operator $\sqrt{-1}$ that signifies the rotation of 90° between τ' and τ'' and γ' and γ'' . The shear modulus can also be represented by a complex variable, ie, the complex dynamic modulus G^* , which is the ratio of the complex stress and complex strain: $G^* = \tau^* / \gamma^*$. The dynamic modulus can also be resolved into two components or vectors (G' and G'') as shown in Figure 18: $G^* = G' + iG''$, where equation 15 holds, and where $G' = G^* \cos \delta$, and $G'' = G^* \sin \delta$.

$$G^* = \left((G')^2 + (G'')^2 \right)^{1/2} \tag{15}$$

The parameter G' is called the storage modulus. It is in phase with the real components of γ^* and τ^* and is a measure of elasticity. It is associated with the energy stored in elastic deformation and is approximately equal to the elastic modulus determined in creep and stress–relaxation experiments when measured at appropriate time scales, ie, $G'(\omega)G(t)$ when $t \rightarrow \infty$. The value of G' is high when a polymer is in its glassy state, but drops with increasing temperature as the polymer goes through the glass transition and becomes soft and rubbery (Fig. 19). If the polymer is cross-linked, the storage modulus does not drop so far after the glass transition. The exact level depends on the degree of cross-linking. For viscoelastic melts it is common practice to associate G' with the ability of a melt to recover from a deformation. However, this has been shown to be invalid in some cases, and an association with the stiffness of the melt is preferred (147).

G'' is called the loss modulus. It arises from the out-of-phase components of γ^* and τ^* and is associated with viscous energy dissipation, ie, damping. The ratio of G'' and G' gives another measure of damping, the dissipation factor or loss tangent (often just called $\tan \delta$), which is the ratio of energy dissipated to energy stored (eq. 16).

$$\tan \delta = G'' / G' \tag{16}$$

Plots of loss modulus or $\tan \delta$ vs temperature for polymers give peaks at energy absorbing transitions, such as the glass transition and low temperature secondary transitions (Fig. 20). Such plots are useful for characterizing polymers and products made from them.

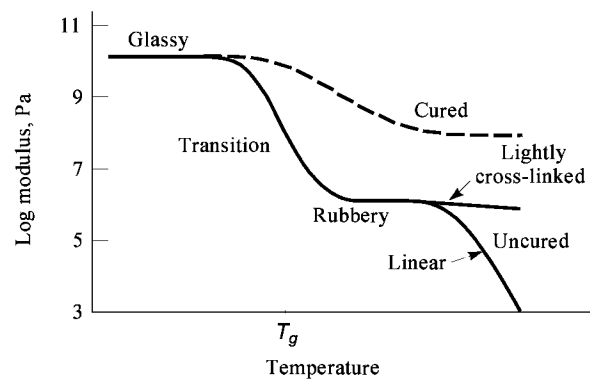


Fig. 19. Generalized modulus–temperature curves for polymeric materials showing the high modulus glassy state, glass-transition regions for cured and uncured polymers, plateau regions for cross-linked polymers, and the dropoff in modulus for a linear polymer.

A viscoelastic material also possesses a complex dynamic viscosity, $\eta^* = \eta' + i\eta''$, and it can be shown that $\eta^* = G^* / i\omega$; $\eta' = G' / \omega$; and $\eta'' = G'' / \omega$, where ω is the angular frequency. The parameter η^* is useful for many viscoelastic fluids in that a plot of its absolute value η^* vs angular frequency in radians/s is often numerically similar to a plot of shear viscosity η vs shear rate. This correspondence is known as the Cox-Merz empirical relationship. The parameter η^* is called the dynamic viscosity and is related to G'' , the loss modulus; the parameter η'' does not deal with viscosity, but is a measure of elasticity.

The significance of G'' , G' , $\tan \delta$, η' , and η'' is that they can be determined experimentally and used to characterize real materials. These parameters depend on frequency and temperature, and this dependence can be used to define behavior. For example, viscoelastic fluids are often characterized by log-log plots of one or more of these quantities vs the angular frequency ω , as shown in Figure 21, which illustrates the behavior of a polymer melt (149).

Normal Stress (Weissenberg Effect). Many viscoelastic fluids flow in a direction normal (perpendicular) to the direction of shear stress in steady-state shear (21,90). Examples of the effect include flour dough climbing up a beater, polymer solutions climbing up the inner cylinder in a concentric cylinder viscometer, and paints forcing apart the cone and plate of a cone–plate viscometer. The normal stress effect has been put to practical use in certain screwless extruders designed in a cone–plate or plate–plate configuration, where the polymer enters at the periphery and exits at the axis.

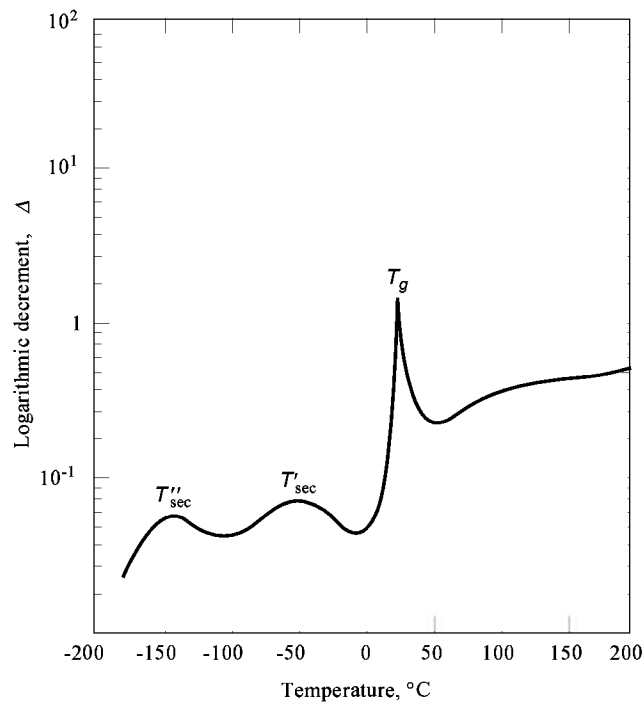


Fig. 20. Logarithmic decrement (related to $\tan \delta$ and loss modulus) vs temperature for a fluorocarbon dibenzoxazole (148). After drying up to 200°C, the experiment was conducted at 200 → 180 → 200°C : $\Delta T / \Delta t = \pm 1.5^\circ\text{C} / \text{min}$ in a helium atmosphere. The T_g gives a sharp damping peak, whereas the secondary glassy state transitions, T_{sec} , are very broad.

The two normal stress functions, $N_1(\dot{\gamma})$ and $N_2(\dot{\gamma})$, are referred to as the first and second normal stress differences, respectively. The former is positive and increases with increasing shear rate, as shown in Figure 22 (149), which describes the steady-shear behavior of the polymer melt in Figure 21. The stress function $N_2(\dot{\gamma})$ is smaller in absolute value than $N_1(\dot{\gamma})$ and is sometimes negative. The first normal stress difference is a useful quantity, as it often gives a good quantitative measure of viscoelasticity. It can be determined from the normal force, which is measurable with several commercial rotational viscometers. In highly elastic liquids it is common for $N_1(\dot{\gamma})$ to be considerably larger than the shear stress.

Description of normal stress measurements on a practical but complex material, paint, is available (150). More recent publications (151–154) give the results of investigations of normal stress differences for a variety of materials. These papers and their references form a useful introduction to the measurement of normal stress differences.

Viscometers

To solve a flow problem or characterize a given fluid, an instrument must be carefully selected. Many commercial viscometers are available with a variety of geometries for wide viscosity ranges and shear rates (10,21,49). Rarely is it necessary to construct an instrument. However, in choosing a commercial viscometer a number of criteria must be considered. Of great importance is the nature of the material to be tested, its viscosity, its elasticity, the temperature dependence of its viscosity, and other variables. The degree of accuracy and precision required, and whether the measurements are for quality control or research, must be considered. The viscometer must be matched to the materials and processes of interest; otherwise, the results may be misleading.

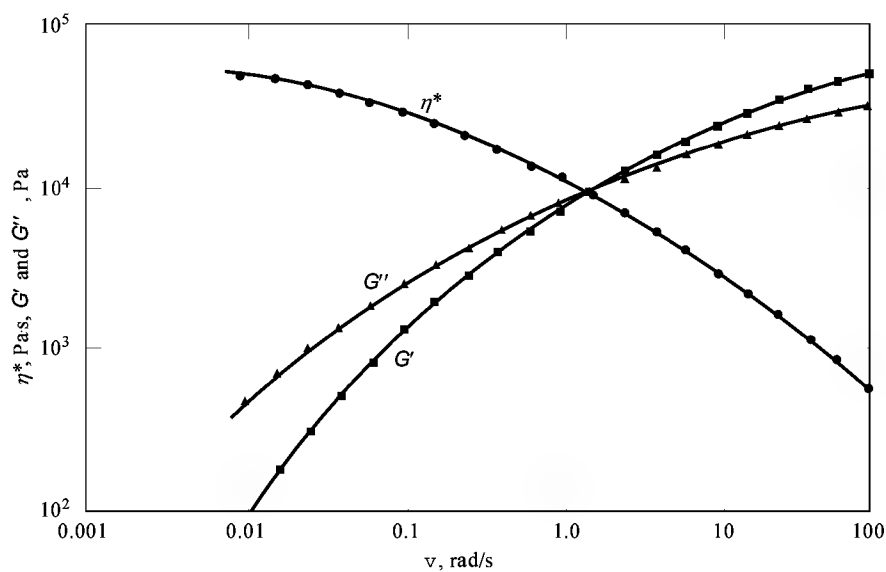


Fig. 21. Dynamic viscoelastic properties of a low density polyethylene (LDPE) at 150°C: complex dynamic viscosity η^* , storage modulus G' , and loss modulus G'' vs angular velocity, ω . To convert Pa·s to P, multiply by 10; to convert Pa to dyn cm², multiply by 10.

Courtesy of Rheometric Scientific.

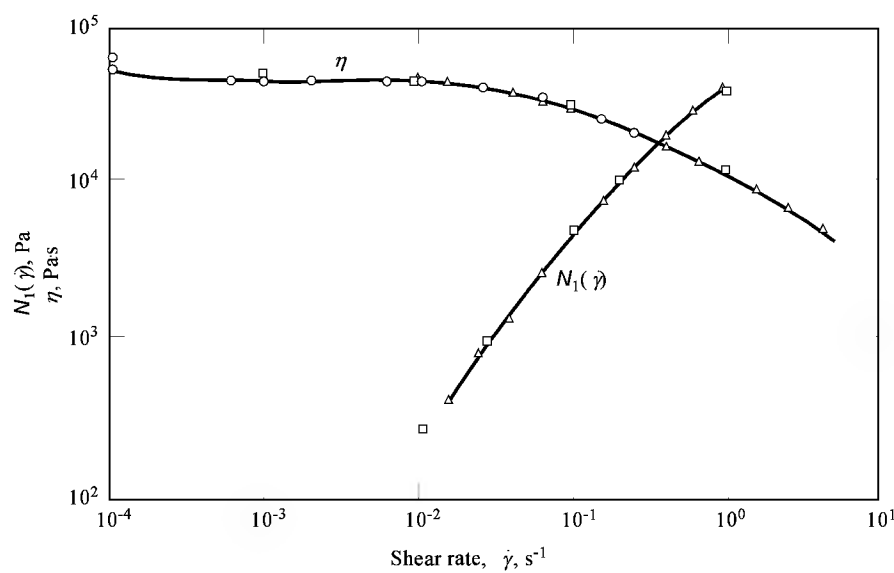


Fig. 22. Shear viscosity η and first normal stress difference $N_1(\dot{\gamma})$ vs shear rate $\dot{\gamma}$ for a low density polyethylene at 150°C (149), where (○) = parallel plate; (△) = cone and plate; and (□) = Weissenberg rheogoniometer. To convert Pa to dyn cm², multiply by 10; to convert Pa·s to P, multiply by 10.

Courtesy of Rheometric Scientific.

Most early viscometers, many of which have been incorporated into standard industrial tests, give single-point measurements. Instead of describing viscosity or shear stress over a range of shear rates, only a single point on the flow curve is produced; the rest of the curve is unknown. This is not a problem with Newtonian liquids because viscosity is independent of shear rate, but it can be misleading in the case of non-Newtonian materials, as shown in Figure 23, where the viscosity profiles of two fluids cross each other. Measurements carried out at the shear rate corresponding to the intersection would indicate that the two materials are identical, even though a simple examination, such as shaking or pouring, would show that they are not. A non-Newtonian fluid cannot be adequately characterized by a single-point measurement, and multipoint measurement techniques are strongly recommended.

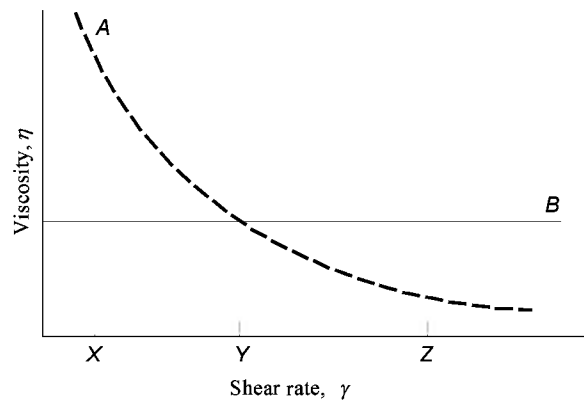


Fig. 23. Viscosity vs shear rate curves for two fluids showing the fallacy of a single point measurement. Fluid *A* would appear to be more viscous than fluid *B* if measured only at point *X*, of the same viscosity if measured at point *Y*, and less viscous if measured only at point *Z*.

Temperature control is also important. Viscosity is highly dependent on temperature; accurate, precise measurements can be made only if the temperature is carefully controlled. More errors are made and more disagreements over viscosity results arise because of incorrect or drifting temperature than for any other reason. Good temperature control can be achieved with commercial baths or circulators equipped with thermostats. Unfortunately, this is impossible with some viscometers. Others lose temperature control because of heat generation at high shear rates, especially with high viscosity materials. Such a temperature increase causes an apparent loss of viscosity under shear. A natural but erroneous conclusion would be that the material is thixotropic or shear thinning.

Viscometers may be separated into three main types: capillary, rotational, and moving body. There are other kinds, usually designed for special applications. For any given type there usually is a choice of several different instruments. The choice depends on the particular requirements of the investigator and the price range.

Capillary Viscometers. Capillary flow measurement is a popular method for measuring viscosity (21,145,146); it is also the oldest. A liquid drains or is forced through a fine-bore tube, and the viscosity is determined from the measured flow, applied pressure, and tube dimensions. The basic equation is the Hagen-Poiseuille expression (eq. 17), where η is the viscosity, r the radius of the capillary, Δp the pressure drop through the capillary, V the volume of liquid that flows in time t , and L the length of the capillary.

$$\eta = \frac{\pi r^4 \Delta p t}{8 V L}$$

(17)

Steady-state, laminar, isothermal flow is assumed. For a given viscometer with similar fluids and a constant pressure drop, the equation reduces to $\eta = K t$ or, more commonly, $\nu = \eta / \rho = C t$, where ρ is the density, ν the kinematic viscosity, and C a constant. Therefore, viscosity can be determined by multiplying the efflux time by a suitable constant.

Capillary viscometers are useful for measuring precise viscosities of a large number of fluids, ranging from dilute polymer solutions to polymer melts. Shear rates vary widely and depend on the instruments and the liquid being studied. The shear rate at the capillary wall for a Newtonian fluid may be calculated from equation 18, where Q is the volumetric flow rate and r the radius of the capillary; the shear stress at the wall is $\tau_w = r \Delta p / 2L$.

$$\dot{\gamma}_w = \frac{4 Q \pi}{r^3}$$

(18)

Absolute viscosities are difficult to measure with capillary viscometers, but viscosities relative to some standard fluid of known viscosity, such as water, are readily determined. The viscometer is calibrated with the reference fluid, and viscosities of other fluids relative to the reference sample are determined from their flow times.

For highly accurate work, corrections must be made for kinetic energy losses, incomplete drainage, turbulence, and possible surface tension and heat effects (21). The largest correction for liquids is that resulting from loss of effective pressure because of the appreciable kinetic energy of the issuing stream. The next most important correction for liquids and the largest for polymer melts is that for energy loss resulting from end effects, ie, viscous resistance caused by velocity gradients as the liquid enters and leaves the capillary (145,146,155). When the terms for these two correction factors are incorporated into the viscosity equation, it is changed to equation 19, where t is the time of flow and B and C are instrument constants generated from measurements of fluids of known viscosity.

$$\nu = \eta / \rho = C t + B / t$$

(19)

These corrections are rarely used in industrial applications because the Poiseuille equation adequately expresses the flow. However, corrections are absolutely necessary for determining even approximate viscosities when using short capillary devices, such as orifice viscometers. Use of a long capillary (>10 dia) and long efflux times (>300 s) minimizes the need for corrections. The ASTM standards D3835 and D5099 describe experimental procedures for using capillary rheometers to measure rheological properties of polymers and rubbers under process conditions.

Glass Capillary Viscometers. The glass capillary viscometer is widely used to measure the viscosity of Newtonian fluids. The driving force is usually the hydrostatic head of the test liquid. Kinematic viscosity is measured directly, and most of the viscometers are limited to low viscosity fluids, ca 0.4–16,000 mm² s. However, external pressure can be applied to many glass viscometers to increase the range of measurement and enable the study of non-Newtonian behavior. Glass capillary viscometers are low shear stress instruments: 1–15 Pa or 10–150 dyn/cm² if operated by gravity only. The rate of shear can be as high as 20,000 s^{–1} based on a 200–800 s efflux time.

The basic design is that of the Ostwald viscometer: a U-tube with two reservoir bulbs separated by a capillary, as shown in Figure 24a. The liquid is added to the viscometer, pulled into the upper reservoir by suction, and then allowed to drain by gravity back into the lower reservoir. The time that it takes for the liquid to pass between two etched marks, one above and one below the upper reservoir, is a measure of the viscosity. In U-tube viscometers, the effective pressure head and therefore the flow time depend on the volume of liquid in the instrument. Hence, the conditions must be the same for each measurement.

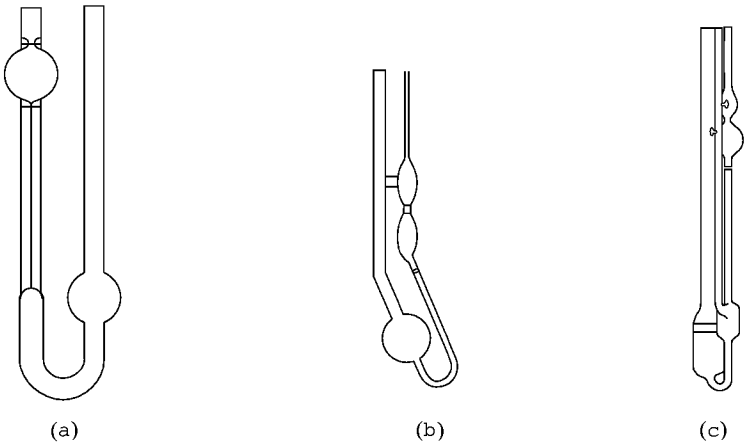


Fig. 24. (a) Ostwald glass capillary viscometer, (b) Cannon-Fenske viscometer, and (c) Ubbelohde viscometer.

The original Ostwald viscometer has been modified in many ways, and a number of different versions are on the market (Table 4) (21). Most are available with a wide choice of capillary diameters and therefore a number of viscosity ranges. A number of viscometers are described in ASTM D445, which also lists detailed recommendations on dimensions and methods of use.

Table 4. Glass Capillary Viscometers^a

Approximate constant, C^b	Viscosity range, $\text{mm}^2 \text{ s}$	Capillary diameters, mm					
		Cannon-Fenske	Ubbelohde	FitzSimons	SIL ^c	Atlantic	Zeitfuchs
0.003	0.6–3.0	0.31 $\pm$ 0.02		0.43 $\pm$ 0.01	0.41 $\pm$ 0.02	0.41–0.42 0.47 $\pm$ 0.01	0.28
0.005	1.0–5.0	0.42					
0.01	2.0–10	0.63	0.58 $\pm$ 0.02	0.61	0.61	0.56	0.38
0.03	6–30	0.78	0.77	0.81 $\pm$ 0.03	0.73	0.74 $\pm$ 0.02	0.50
0.05	10–50		0.87		0.91	0.84	
0.1	20–100	1.02	1.10 $\pm$ 0.03	1.05	1.14 $\pm$ 0.03	1.00	0.67
0.3	60–300	1.26	1.43	1.32 $\pm$ 0.04	1.50	1.31 $\pm$ 0.02	0.88
0.5	100–500	1.48	1.64		1.71	1.48	
1	200–1,000	1.88	1.95	1.96	2.03	1.77 $\pm$ 0.03	1.20
3	600–3,000	2.20 $\pm$ 0.05	2.67 $\pm$ 0.04		2.80	2.34	1.42
5	1,000–5,000	3.10	3.06		3.06	2.65 $\pm$ 0.04	
10	2,000–10,000	4.0	3.62		3.79 $\pm$ 0.04		1.93
30	6,000–30,000						2.52
100	(2–10) $\times 10^4$						3.06

^a Ref. 21.
^b Constant for equation $\nu = C/t$, where ν = kinematic viscosity in $\text{mm}^2 \text{ s}$.
^c Standard Inspection Laboratories.

The Cannon-Fenske viscometer (Fig. 24b) is excellent for general use. A long capillary and small upper reservoir result in a small kinetic energy correction; the large diameter of the lower reservoir minimizes head errors. Because the upper and lower bulbs lie on the same vertical axis, variations in the head are minimal even if the viscometer is used in positions that are not perfectly vertical. A reverse-flow Cannon-Fenske viscometer is used for opaque liquids. In this type of viscometer the liquid flows upward past the timing marks, rather than downward as in the normal direct-flow instrument. Thus the position of the meniscus is not obscured by the film of liquid on the glass wall.

The Ubbelohde viscometer is shown in Figure 24c. It is particularly useful for measurements at several different concentrations, as flow times are not a function of volume, and therefore dilutions can be made in the viscometer. Modifications include the Cannon-Ubbelohde, semimicro, and dilution viscometers. The Ubbelohde viscometer is also called a suspended-level viscometer because the liquid emerging from the lower end of the capillary flows down only the walls of the reservoir directly below it. Therefore, the lower liquid level always coincides with the lower end of the capillary, and the volume initially added to the instrument need not be precisely measured. This also eliminates the temperature correction for glass expansion necessary for Cannon-Fenske viscometers.

For accurate and precise measurement the glass capillary must be clean. The viscometer must be cleaned thoroughly after each series of operations. Samples being tested and cleaning solvents should be filtered to remove particles that can clog the capillary.

All glass capillary viscometers should be calibrated carefully (21). The standard method is to determine the efflux time of distilled water at 20°C. Unfortunately, because of its low viscosity, water can be used only to standardize small capillary instruments. However, a calibrated viscometer can be used to determine the viscosity of a higher viscosity liquid, such as a mineral oil. This oil can then be used to calibrate a viscometer with a larger capillary. Another method is to calibrate directly with two or more certified standard oils differing in viscosity by a factor of approximately five. Such oils are useful for calibrating virtually all types of viscometers. Because viscosity is temperature-dependent, particularly in the case of standard oils, temperature control must be extremely good for accurate calibration.

In recent years several commercial capillary viscometers that allow automation have been developed. For example, the time of flow is measured by the interruption of a light beam focused on a photoelectric cell. Such a device can be attached to an existing viscometer to facilitate measurements and improve precision. Some completely automatic instruments control the temperature, fill the viscometer, take several readings (sometimes with dilutions), as well as clean, rinse, and dry the viscometer before the next filling. The instrument prints efflux times and in some cases can be programmed to print viscosities. Manufacturers of automated viscosity systems include Cannon, Schott, Design Scientific, and Lauda; Wescan supplies timer and photocell combinations. An inexpensive timer is based on light pipes and an electronic stopwatch (156). Cannon Instrument Company supplies a wide range of glass capillary viscometers, temperature baths, viscosity standards, and related equipment.

A further development is the capillary bridge differential viscometer (157), which is based on a fluid analogue of the Wheatstone bridge. This device (Viscotek Corporation) utilizes differential pressure measurement on separate capillary flow streams of solvent and polymer solution. The specific viscosity, η_{sp} , is determined from the differential pressure and the inlet solvent pressure. The technique is highly sensitive, and a limiting viscosity number (intrinsic viscosity) is normally determined from measurements on a single, very dilute polymer solution. Approximately half of these instruments are sold as gel-permeation chromatography (gpc) detectors useful for detecting and measuring polymer branching. Such an instrument has been used for on-line intrinsic viscosity measurements. Refractive index and viscosity are measured at the same time. The former is used to determine the concentration, which is then combined with the viscosity result to give intrinsic viscosity.

Orifice. Orifice viscometers, also called efflux or cup viscometers, are commonly used to measure and control flow properties in the manufacture, processing, and application of inks, paints, adhesives, and lubricating oils. Their design answered the need for simple, easy-to-operate viscometers in areas where precision and accuracy are not particularly important. In these situations knowledge of a true viscosity is unnecessary, and the efflux time of a fixed volume of liquid is a sufficient indication of the fluidity of the material. Examples of orifice viscometers include the Ford, Zahn, and Shell cups used for paints and inks and the Saybolt Universal and Furol instruments used for oils (Table 5).

Orifice viscometers usually have extremely short capillaries. The typical orifice viscometer is a cup with a hole in the bottom. The cup is filled, and the time required for the liquid to flow out is measured. The hydrostatic head decreases as the liquid flows, and there is a large kinetic energy effect. Flow analysis shows that the flow does not follow the Hagen-Poiseuille law, and efflux times are not related to viscosities in any simple manner. Therefore, it is better not to convert efflux times to viscosities except during calibration with standard oils. Efflux times should be accepted as arbitrary measures of viscosity and noted in terms of the viscometer being used, ie, Saybolt seconds, Ford seconds, etc. The precision of orifice viscometers is poor because of the lack of temperature control, wear during use, and variations in manufacture. However, they are widely used in industry because they are inexpensive, robust, and easy to use. Some of these instruments have the added convenience of being capable of dipping into or remaining in the material to be tested. Dip cups determine approximate or relative viscosities in resin reactors, ink reservoirs, paint dip tanks, adhesive mixing tanks, etc. These are applications where the limitations of orifice cups are unimportant.

Orifice viscometers should not be used for setting product specifications, for which better precision is required. Because they are designed for Newtonian and near-Newtonian fluids, they should not be used with thixotropic or highly shear-thinning materials; such fluids should be characterized by using multispeed rotational viscometers.

Some orifice viscometers, such as the Shell dip cup and the European ISO cup, which resembles a Ford cup with a capillary, have long capillaries. These cups need smaller kinetic energy corrections and give better precision than the corresponding short-capillary viscometers. However, they are still not precision instruments, and should be used only for control purposes.

Table 5. Some Orifice Viscometers Used in Industry

Viscometer ^a	Orifice length, mm	Orifice diameter, cm	Viscosity range, mm ² s(cSt)	Approximate constants ^b		Main applications
				<i>k</i>	<i>K</i>	
DIN ^c No. 4		0.40	60–460	4.67	570	paints, varnishes
Fisher ^d No. 2		0.4	35–230	2.32	200	paints
Ford ^{e,f} , No. 2		0.25	30–120	1.30	100	paints, varnishes
				(for 40 < <i>t</i> < 100 s)	0	
No. 3		0.34	50–250	2.30	800	paints, varnishes
				(for 30 < <i>t</i> < 100)		
No. 4		0.41	70–400	3.70	400	paints, varnishes
ISO ^g						
No. 3	20 ± 0.05	0.30	10–45	0.443	200	paints, inks
No. 4	20 ± 0.05	0.40	30–160	1.37	200	paints, inks
No. 6	20 ± 0.05	0.60	200–700	6.90	570	paints, inks
Redwood ^h (U.K.) No. 1	10 ± 0.05	1.62 min	1–500	0.264	190	petroleum products
				(for 40 < <i>t</i> < 85 s)		
				0.247		
				(for 85 < <i>t</i> < 200 s)		
Redwood ^h No. 2	50 ± 0.2	3.8	500–5000	2.6	40	petroleum products
Saybolt Universal ^h	12.2	1.76	1–400	0.226	195	petroleum products
				(for <i>t</i> < 100 s)		
Saybolt Furol ^h	12.2	3.15	400–4000	0.220	130	petroleum products
				(for <i>t</i> < 100 s)		
Shell ^{f,1}						
No. 1	25	0.18	4–20	0.19	48	inks
No. 2	25	0.24	10–60	0.55	46	inks
No. 3	25	0.31	30–150	1.50	37	paints, inks
No. 4	25	0.38	70–300	3.45	14	paints, inks
No. 5	25	0.46	125–500	6.5	42	inks
No. 6	25	0.58	320–1500	16.2	34	
Zahn ^{e,f}						
No. 1		0.20	20–100	0.80	400	varnishes
				(for 40 < <i>t</i> < 100 s)		
No. 2		0.27	50–300	2.93	500	paints, inks
No. 3		0.38	100–1000	10.7	180	paints, inks
					0	
No. 4		0.43	200–1500	14.0	120	paints
					0	
No. 5		0.53	400–2000	23.3	200	bodied paints

^b Constants *k* and *K* are based on $\nu = kt - K/t$, in mm² s(cSt). Unless otherwise noted, constants hold for 20 < *t* < 100 s.

^a No thermostat; viscosity unit – s, unless otherwise noted.

^c DIN (Deutsche Normen) cup constants from DIN Standard 53211.

^d Fisher cup constants based on data from only two cups.

^e Ford and Zahn cup constants (23).

^f Linear equations for Ford, Shell, and Zahn cups (158).

^g ISO cup constants are from ISO International Standard 2431.

^h Thermostatted, viscosity units – s 60 mL.

¹ Based on data from manufacturers of Shell cups (158) (NORCROSS Corp.).

If it is necessary to calculate kinematic viscosities from efflux times, such as in a calibration procedure, equation 20 should be used, where *t* is the efflux time and *k* and *K* are constants characteristic of the particular viscosity cup (see Table 5) (158,159).

$$\nu = kt - K/t$$

(20)

In most cases it is sufficient to be able to convert from one viscometer value to another or to approximate kinematic viscosities with the help of charts or tables; literature from manufacturers is useful.

Linear equations of the type $\nu = \epsilon t - C/t$, where ϵ and *C* are constants, relate kinematic viscosity to efflux time over limited time ranges. This is based on the fact that, for many viscometers, portions of the viscosity–time curves can be taken as straight lines over moderate time ranges. Linear equations, which are simpler to use in determining and applying correction factors after calibration, must be applied carefully as they do not represent the true viscosity–time relation. Linear equation constants have been given (158) and are used in ASTM D4212.

Piston Cylinder (Extrusion). Pressure-driven piston cylinder capillary viscometers, ie, extrusion rheometers (Fig. 25), are used primarily to measure the melt viscosity of polymers and other viscous materials (21,47,49,50). A reservoir is connected to a capillary tube, and molten polymer or another material is extruded through the capillary by means of a piston to which a constant force is applied. Viscosity can be determined from the volumetric flow rate and the pressure drop along the capillary. The basic method and test conditions for a number of thermoplastics are described in ASTM D1238. Melt viscoelasticity can influence the results (160).

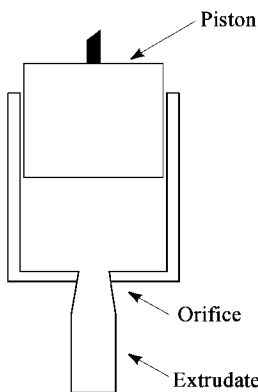


Fig. 25. Piston cylinder capillary viscometer (49).

Polymer melts are frequently non-Newtonian. In this case the earlier expression given for the shear rate at the capillary wall does not hold. A correction factor $(3n + 1) / 4n$, called the Rabinowitsch correction, must be applied in such a way that equation 21 applies, where $\dot{\gamma}_{tw}$ is the true shear rate at the wall and n is a power law factor (eq. 22) determined from the slope of a log–log plot of the true shear stress at the wall, τ_{tw} , vs $\dot{\gamma}_{tw}$. For a Newtonian liquid, $n = 1$. A true apparent viscosity, η_t , can be calculated from equation 23.

$$\dot{\gamma}_{tw} = \frac{(3n + 1)}{4n} \dot{\gamma}_w$$

(21)

$$n = \frac{d \log \tau_{tw}}{d \log \dot{\gamma}_w}$$

(22)

$$\eta_t = \frac{\tau_{tw}}{\dot{\gamma}_{tw}}$$

(23)

An even more important correction for polymer melts is that for end effects. Entrance and exit effects cause pressure drops that interfere with an accurate determination of the pressure gradient, $\Delta p / L$, needed for viscosity determination using the Hagen–Poiseuille equation. It has been assumed (155) that the effective length of the capillary is greater than the actual length, and the shear stress at the wall has been corrected as equation 24, where R is the capillary radius, L is the actual length, and E is an empirical parameter obtained by extrapolating a plot of Δp versus the capillary aspect ratio, L / R , to zero pressure drop at constant shear rate, $\dot{\gamma}$, for capillaries of different length, as shown in Figure 26 (145).

$$\tau_w^c = \frac{R \Delta p}{2(L + ER)}$$

(24)

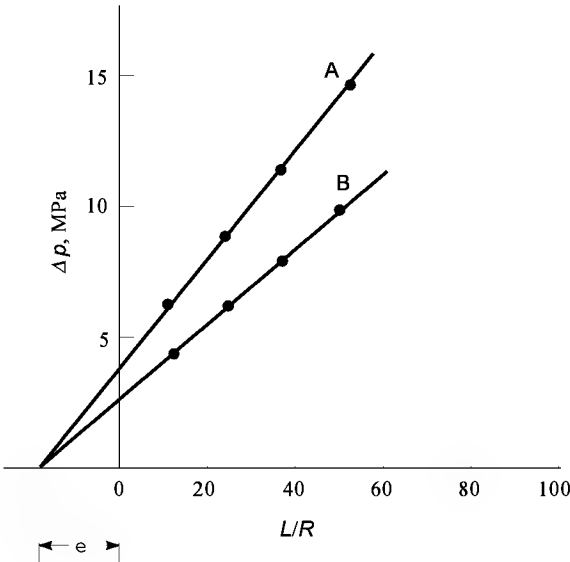


Fig. 26. Bagley plot of pressure, Δp , drop along a capillary versus capillary aspect ratio, L / R , at A, $\dot{\gamma} = 590 \text{ s}^{-1}$, and B, $\dot{\gamma} = 295 \text{ s}^{-1}$. To convert MPa to psi, multiply by 145.

A number of instruments are based on the extrusion principle, including slit flow and normal capillary flow (Table 6). These instruments are useful when large numbers of quality control or other melt viscosity test measurements are needed for batches of a single material or similar materials. When melt viscosities of a wide range of materials must be measured, rotational viscometers are preferable. Extrusion rheometers have been applied to other materials with some success with adhesives and coatings (10,161).

Table 6. Selected Extrusion Viscometers^a

Manufacturer	Model name and number	Drive mechanism	Price range ^b
Burrell	Burrell-Severs A-250	gas pressure	A
Custom Scientific	CS-127 HS	gas pressure	A

Davenport	High Shear Viscometer	gas pressure	A
	Extrusion Rheometer	electromechanical	B
Gnttferf	Rheograph 2002	hydraulic	D
	Rheo Tester		C
	RTR		
	Melt Index Tester		
Instron	3210 Accessory	electromechanical	B
	3211 Capillary Rheometry	electromechanical	C
Kayeness (Dynisco)	Galaxy III, V Automated		
	Capillary Rheometers		
	Galaxy I		
	Melt Index Tester		
MTS	MTS Capillary Viscometer	servohydraulic	D
Monsanto	Automatic Capillary Rheometer	gas pressure, pneumatic cylinder	B
	Processibility Tester		C
Paar	HVA 6 Automated Capillary Slit Viscometer	gas pressure	
Rheometrics	MFM Melt Flow Rheometer		
Rosand	Automatic Melt Flow Indexer		D
Seiscor	AK Continuous Capillary Rheometer	gear pump	B
	Han Slit Rheometer	gear pump	B
Shimadzu	CFT-500	electromechanical	C
Tinius Olsen	Sieglaﬀ-McKelvey	gas pressure, pneumatic cylinder	C
Toyo Seiki	Westover High Pressure	gas pressure, hydraulic	B

^a From Ref. 49 and manufacturers' literature.
^b 1995 Price ranges: A = \$8,000–\$12,000; B = \$12,000–\$30,000; C = \$30,000–\$60,000; D >\$60,000.

Rotational Viscometers. Rotational viscometers consist of two basic parts separated by the fluid being tested (21,23,49,145,146,158,162–166). The parts may be concentric cylinders (cup and bob), plates, a low angle cone and a plate, or a disk, paddle, or rotor in a cylinder. Rotation of one part against the other produces a shearing action on the fluid. The torque required to produce a given angular velocity or the angular velocity resulting from a given torque is a measure of the viscosity. Rotational viscometers are more versatile than capillary viscometers. They can be used with a wide range of materials because opacity, settling, and non-Newtonian behavior do not cause difficulties. Viscosities over a range of shear rates and as a function of time can be measured. Therefore, they are useful for characterizing shear thinning and time-dependent behavior. Rotational viscometers are considerably more complicated mechanically than most capillary viscometers. They range from relatively simple, inexpensive (<\$5,000), low shear devices, eg, the Brookfield and the Stormer, to complicated and expensive (ca \$100,000) instruments capable of steady-state and oscillatory motion, eg, the Weissenberg rheogoniometer and Rheometrics' fluids rheometer.

In the 1990s rotational viscometers have been developed that are either integral or closely interfaced with computers for operation and control of the instrument as well as for data collection, reduction, and storage. The combination of a computer, modern electronics, and feedback loops provides great flexibility and control. For example, pulses from the computer coupled with feedback from the viscometer can give precise oscillatory motion as well as shear viscosity measurement at constant stress or speed; only simple hardware is involved. Such instruments are useful for rheological measurements and studying the structure of dispersions and formulated products. These new instruments are versatile, easy to use, and allow the collection, analysis, and comparison of large amounts of data in a short time.

The mechanical parts of computer-controlled viscometers can be simple and should not become obsolete for many years. The complexity is in the software, and this is where changes will have to be made to keep the instrument up to date. Some manufacturers are offering new, updated software free for a period of several years after the instrument is purchased.

Rotational viscometers often were not considered for highly accurate measurements because of problems with gap and end effects. However, corrections can be made, and very accurate measurements are possible. Operating under steady-state conditions, they can closely approximate industrial process conditions such as stirring, dispersing, pumping, and metering. They are widely used for routine evaluations and quality control measurements. The commercial instruments are effective over a wide range of viscosities and shear rates (Table 7).

The equations and methods for determining viscosity vary greatly with the type of instrument, but in many cases calculations may be greatly simplified by calibration of the viscometer with a standard fluid, the viscosity of which is known for the conditions involved. General procedures for calibration measurement are given in ASTM D2196. The constant thus obtained is used with stress and shear rate terms to determine viscosity by equation 25, where the stress term may be torque, load, or deflection, and the shear rate may be in rpm, revolutions per second (rps), or s⁻¹.

$$\eta = K(\text{stress term} \div \text{shear rate term})$$

(25)

Constants and factors are often supplied by the manufacturer. Separate constants may be given for converting the stress and shear rate terms to the correct quantities and units.

Table 7. Rotational Viscometers

Viscometer	Price range ^a	Viscosity range ^b , mPa·s (cP)	Shear rate capability, s ⁻¹	Temperature control	Manufacturer
ATO RheoSystems Stresstech	D	0.5–10 ¹²	10 ⁻⁵ –5 × 10 ⁵ (10 ⁻⁶ in creep)	good (to 150°C); higher temperatures optional	Reologica Instrument A.B., Lund, Sweden; ATS RheoSystems, Wrightstown, N.J.
Bohlin					Bohlin Reologi, Lund, Sweden; Cranbury, N.J.
VOR	D	10 ⁻¹ –10 ¹¹	10 ⁻¹ –10 ⁴	good	
VOR-M	D		10 ⁻⁶ –10 ³	good (to 550°C)	
CS	D			good	
CS-M	D			good (to 400°C)	
Visco 88 BV	B	6.5–3.5 × 10 ⁵	4–1,200		
Brabender Plasti-Corder	C	1–10 ⁵	low shear rate	fair to good	C. W. Brabender OHG, Duisburg, Germany
Brookfield Basic Viscometer	A	1–10 ⁰	~0.1–150	none to good,	Brookfield Engineering Laboratories, Stoughton, Mass.

				depending on system used	
cone plate	A	5–10 ⁰	<1500	good	
CAP 1000	A	25–10,000	3,000; 12,000	good	
CAP 2000	A–B	10–1.5 × 10 ⁵	166–26,600	good	
KU-1 Stormer-type	A	100–5,000	~50	external control, usually poor	
Coesfeld					Coesfeld Material Test GmbH & Co., Dortmund, Germany
Visco-mix	C	10–10 ⁸		external control	
Rheosyst	C	1–10 ⁹			
Haake					Haake GmbH, Karlsruhe, Germany; Paramus, N.J.
RV 20	D	1–10 ⁰	4 × 10 ^{–4} –4 × 10 ⁴	good	
RheoStress RS 100	D	1–10 ⁰	10 ^{–6} –10 ³	good (to 500°C)	
Viscotester VT 550	A	0.5–10 ⁷	0.1–10 ⁴	external control	
Hercules	B	1–10	<115,000	poor	Kaltec Scientific Instrument, Kalamazoo, Mich.
ICI Cone Plate	A	5–5 × 10 ⁴	3,000; 12,000	good	Research Equipment Ltd., Hampton, Middlesex, U.K.
MacMichael	A	<10 ⁶	2–12	fair	Fisher Scientific, Pittsburgh, Pa.
Mooney Disk	A	<10 ⁸	~1	fair	Monsanto Co., Akron, Ohio
Nametre Rotary	B	5–5 × 10 ⁴	<200	none to good	Nametre Co., Edison, N.J.
Physica					Paar Physica, Stuttgart, Germany; Edison, N.J.
Viscolab	C–E	1–10 ¹²	10–10 ⁵	good	
Rheolab	C–E	1–10 ¹²	10–10 ⁵	good	
Ravenfield BS	B	10 ^{–2} –10 ⁸		good	Ravenfield Designs, Ltd., Heywood, Lancashire, U.K.
Rheometric					Rheometric Scientific, Inc., Piscataway, N.J.
RFS II Fluids Rheometer	D–E	10–10 ⁶	10 ^{–1} –10 ⁴	good	
RSR Controlled Stress Rheometer	D–E	10 ^{–1} –10 ⁸		good	
Rheo-Tech Viscoelastic Rheometer	C–D	0.5–5 × 10 ⁸	10 ^{–6} –10 ⁴	good	Rheo-Tech International, London, U.K.
Stormer (Standard and Digital)	A	100–10 ⁴	~50	none to fair	Cannon Instrument, State College, Pa. Thomas Scientific, Swedesboro, N.J.
TA Instruments					TA Instruments, New Castle, Del.
CSL ² Controlled Stress Rheometer	D–E	10 ^{–2} –10 ⁰	10 ^{–6} –5 × 10 ³	good	
Weissenberg Rheogoniometer	E	10 ^{–1} –10 ¹⁰	10 ^{–4} –10 ⁴	good	

^a 1995 price ranges: A = <\$10,000; B = \$10,000–\$25,000; C = \$25,000–\$50,000; D = \$50,000–\$100,000; and E >\$100,000.
^b Values are approximate.

A constant is often determined from measurements with a Newtonian oil, particularly when the calibrations are supplied by the manufacturer. This constant is valid only for Newtonian specimens; if used with non-Newtonian fluids, it gives a viscosity based on an inaccurate shear rate. However, for relative measurements this value can be useful. Employment of an instrument constant can save a great deal of time and effort and increase accuracy because end and edge effects, slippage, turbulent interferences, etc, are included.

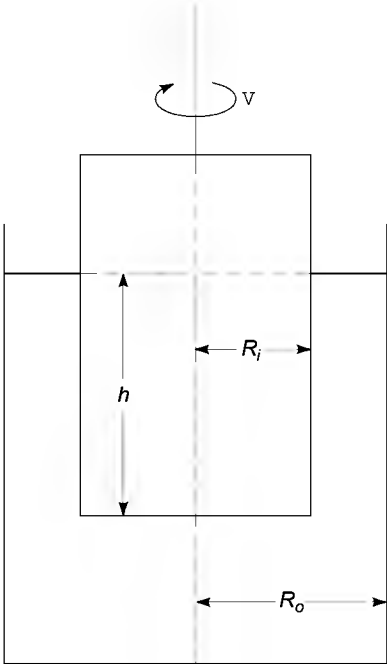


Fig. 27. Concentric cylinder viscometer. R_i and R_o are the radii of the inner and outer cylinder, respectively, and Ω is the relative angular velocity.

Coaxial (Concentric Cylinder) Viscometer. The earliest and most common type of rotational viscometer is the coaxial or concentric cylinder instrument. It consists of two cylinders, one within the other (cup and bob), keeping the specimen between them, as shown in Figure 27. The first practical rotational viscometer consisted of a rotating cup with an inner cylinder supported by a torsion wire. In variations of this design the inner cylinder rotates. Instruments of both types are useful for a variety of applications.

The relationship between viscosity, angular velocity, and torque for a Newtonian fluid in a concentric cylinder viscometer is given by the Margules equation (eq. 26) (21,146), where M is the torque on the inner cylinder, h the length of the inner cylinder, Ω the relative angular velocity of the cylinder in radians per second, R_i the radius of the inner cylinder wall, R_o the radius of the outer cylinder wall, and k an instrument constant.

$$\eta \left(\frac{M}{\Omega 4 \pi h} \right) \left(\frac{1}{R_i^2} - \frac{1}{R_o^2} \right) = \frac{k M}{\Omega} \tag{26}$$

Therefore, the viscosity can be determined from the torque and angular velocity. However, the viscosity is usually calculated from the shear rate and shear stress, which can be obtained from the Margules equation. The shear rate is given by equation 27, where r is any given radius.

$$\dot{\gamma} = \frac{(2 \Omega \ r^2) \left(\frac{R_i^2 R_o^2}{R_o^2 - R_i^2} \right)}{\left(\frac{R_o^2}{R_i^2} - \frac{R_i^2}{R_o^2} \right)} \tag{27}$$

The shear stress is given by equation 28:

$$\tau = \frac{M}{2 \pi r^2 h} \tag{28}$$

The shear rate and shear stress can be calculated for any radius r from these equations. In most cases the radius used is R_i because the shear stress and shear rate of interest are at the inner, torque-sensing cylinder. Thus equations 27 and 28 become

$$\dot{\gamma} = \frac{2 \Omega R_o^2}{(R_o^2 - R_i^2)} \quad \text{and} \quad \tau = \frac{M}{2 \pi R_i^2 h}$$

The viscosity of a Newtonian fluid may be determined from the Margules equation or from the slope of a shear stress–shear rate plot. Non-Newtonian fluids give intercepts and curves with such plots. Viscosities can be calculated, but accurate values depend on including correction factors for yield points and shear thinning, ie, shear rate corrections, in the above equations (21). The shear rate correction may be minimized by using a very small gap size; in other words, the ratio of the inner to outer radius should be as close to unity as possible. In practical terms this means maintaining a ratio of 0.95, which is impossible with many sensors used with commercial rotational viscometers. With highly shear thinning materials even this is insufficient, and corrections must be made regardless of the gap size. However, with viscosity–shear rate curves, although correction shifts each point by a rather large amount, the corrected curve itself is only slightly different from the uncorrected curve (167).

In addition to non-Newtonian flow, the main correction necessary for concentric cylinder measurements is that on account of end effects. Because the inner cylinder is not infinitely long, there is drag on the ends as well as on the face of the cylinder. The correction appears as an addition, h_o , to the length, h . The correction is best determined by measuring the angular velocity and torque at several values of h , that is, at various depths of immersion. The data are plotted as M / Ω vs h , and extrapolation is made to a value of h_o at $M / \Omega = 0$. The quantity $(h + h_o)$ is substituted for h in the various equations.

Cone–Plate Viscometer. In a cone–plate viscometer (Fig. 28), a low angle ($<3^\circ$) cone rotates against a flat plate with the fluid sample between them. The cone–plate instrument is a simple, straightforward device that is easy to use and extremely easy to clean. It is well suited to routine work because measurements are rapid and no tedious calculations are necessary. With careful calibration and good temperature control it can also be used for research. Heated instruments can be used for melt viscosity measurements.

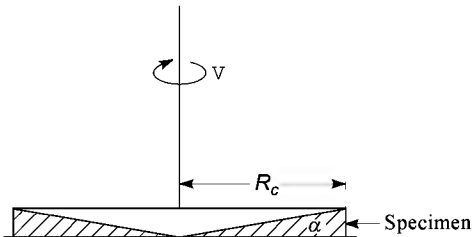


Fig. 28. Cone–plate viscometer. R_c is the radius of the cone, α is the angle between cone and plate, and Ω is the relative angular velocity.

In most rotational viscometers the rate of shear varies with the distance from a wall or the axis of rotation. However, in a cone–plate viscometer the rate of shear across the conical gap is essentially constant because the linear velocity and the gap between the cone and the plate both increase with increasing distance from the axis. No tedious correction calculations are required for non-Newtonian fluids. The relevant equations for viscosity, shear stress, and shear rate at small angles α of Newtonian fluids are equations 29, 30, and 31, respectively, where M is the torque, R_c the radius of the cone, v the linear velocity, and r the distance from the axis.

$$\eta = \frac{3 \alpha M}{2 R_c^3} \tag{29}$$

$$\tau = \frac{3 M}{2 \pi R_c^3} \tag{30}$$

$$\dot{\gamma} = \frac{dv}{dr} = \frac{\Omega}{\alpha} \tag{31}$$

Cone–plate geometry has several advantages over concentric cylinder geometry, including a smaller sample size, a homogeneous shear rate, and easy conversion of data. Disadvantages include the need for precise adjustment of the gap, including resetting when the temperature is changed, specimen drying, solvent evaporation, slinging of material from the gap, and the possibility of viscous heating, particularly at high shear rates. The last problem is compounded by the fact that temperature control with commercial instruments is not always as good as it should be.

Parallel Plate Viscometer. In parallel plate viscometers (164) the gap width usually is larger and can be varied freely. This is an advantage when measuring suspensions or dispersions with large particles or with a tendency to fly out of the gap. The wide gap means that there is less sensitivity to

temperature changes. Therefore, resetting usually is not necessary and temperature scans are much easier to run than with a cone–plate viscometer. However, with the plate–plate instrument, the velocity, and therefore the shear rate, varies with the distance from the center of the plate. This makes viscosity data more difficult to evaluate.

The maximum shear rate (at the plate rim) is given by equation 32, where R_p is the radius of the plate and h the distance between the two plates.

$$\dot{\gamma}_m = \frac{\Omega R_p}{h} \tag{32}$$

The viscosity is given by equation 33, where M is the torque.

$$\eta = \frac{3M}{2\pi R_p^3 \dot{\gamma}_m} \left[1 + 3 \frac{d \ln M}{d \ln \dot{\gamma}} \right] \tag{33}$$

Dynamic Viscometer. A dynamic viscometer is a special type of rotational viscometer used for characterizing viscoelastic fluids. It measures elastic as well as viscous behavior by determining the response to both steady-state and oscillatory shear. The geometry may be cone–plate, parallel plates, or concentric cylinders; parallel plates have several advantages, as noted above.

Controlled Stress Viscometer. Most rotational viscometers operate by controlling the rotational speed and, therefore, the shear rate. The shear stress varies uncontrollably as the viscosity changes. Often, before the structure is determined by viscosity measurement, it is destroyed by the shearing action. Yield behavior is difficult to measure. In addition, many flow processes, such as flow under gravity, settling, and film leveling, are stress-driven rather than rate-driven.

An instrument where the shear stress and rate of stress change, rather than the rotational speed, are controlled offers advantages. A few such instruments have existed for many years, eg, the Stormer; others have been developed more recently (168–173). A typical instrument consists of a drag cup motor, a frictionless air-bearing torque shaft, sensors for measuring angular deflection and velocity, and a rotating bob and fixed cup or parallel plates (170–172). The rotating shaft must be suspended in a frictionless manner to permit measurements at very low stresses.

Controlled stress viscometers are useful for determining the presence and the value of a yield stress. The structure can be established from creep measurements, and the elasticity from the amount of recovery after creep. The viscosity can be determined at very low shear rates, often in a Newtonian region. This zero-shear viscosity, η_0 , is related directly to the molecular weight of polymer melts and concentrated polymer solutions.

Yield stresses are determined with the help of a vane cylinder acting as sensor (174), which is attached to a stress rheometer (175). The sensor has four to eight thin blades centered around the cylindrical shaft. The instrument is immersed in the material, and a stress is applied. If the stress is below the yield stress, the material deforms elastically (small strain), but does not yield. When the applied stress exceeds the yield stress, the material flows and the sensor rotates continuously. Usually, a series of measurements are carried out at successively higher stresses, as shown in Figure 29 (175). The yield behavior is seen as a large strain over a short period of time. The torque for yielding is between the last nonyield torque value and the one that produced continuous rotation. Subsequent tests are performed with smaller torque increments to narrow the range for the yield stress. Details of the method and the equation relating yield stress to torque are available (175). Controlled stress techniques are used to determine yield stresses and other flow properties for polymer melts (54) as well as for solutions and dispersions.

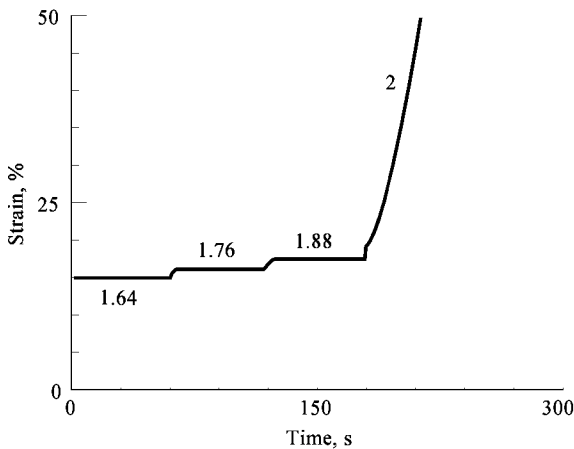


Fig. 29. Measurements of yield stress with a vane device and Rheometrics controlled stress rheometer. The torque required to cause yielding is between 1.88 and 2mN·m as indicated on the curve. To convert mN·m to dyn·cm, multiply by 10⁴.

Courtesy of the Society of Rheology.

Other Rotational Viscometers. Some rotational viscometers employ a disk as the inner member or bob, eg, the Brookfield and Mooney viscometers; others use paddles (a geometry of the Stormer). These nonstandard geometries are difficult to analyze, particularly for an infinite bath, as is usually employed with the Brookfield and the Stormer. The Brookfield disk has been analyzed for Newtonian and non-Newtonian fluids and shear rate corrections have been developed (22). Other nonstandard geometries are best handled by determining instrument constants by calibration with standard fluids.

Another type of rotational viscometer is the helical-screw rheometer (176). This instrument is basically a screw-type metering pump that does not pump. The measure of force is the pressure difference resulting from the rotational motion. It is possible to use a bank of pressure transducers of different sensitivities to measure viscosity over a wide range. The instrument can be used for high temperature rheometry and to follow polymerization, shear and heat degradation, and other developments.

Specific Commercial Rotational Viscometers. Information on selected commercial rotational viscometers can be found in Table 7. The ATS RheoSystems Stresstech rheometer is an instrument that combines controlled stress as well as controlled strain (shear rate) and oscillatory measurements. It has a torque range of 10⁻³ to 50 mN·m, an angular velocity range of 0 to 300 rad/s, and a frequency range of seven decades. Operation and temperature programming (–30 to 150°C; higher temperatures optional) are computer controlled.

The Bohlin rheometer system includes two viscometers (models CVO and VOR) that do both steady shear and oscillatory measurements. Both instruments are computer operated, but CVO has a number of advanced features including automatic gap setting and adjustment, normal force, high speed capability, and a very wide torque range. Both operate in steady shear or shear rate sweep modes. The oscillatory mode is useful for the characterization of viscoelastic materials. The VOR viscometer has a shear rate range of 10⁻¹–10⁴ s⁻¹; the CVO, 10⁻¹–10⁵ s⁻¹. The viscosity range is 10⁻¹–10¹¹ mPa·s. A useful rapid thixotropic recovery mode employs special software for high speed rotation, followed by a small amplitude oscillatory motion. Thixotropic recovery can be followed as a function of time after shearing. The melt rheometer version of the VOR is the VOR-M which has a temperature range of ambient to 550°C at shear rates of 10⁻⁶–10³ s⁻¹ and frequencies of 10⁻³ to 20 Hz.

Bohlin also markets a computer-driven controlled stress rheometer, which determines viscosity as a function of stress and measures creep and

recovery. The standard sensor is a concentric cylinder, but cone-plate and parallel plate geometries are also available. The maximum torque for sustained periods is 10 mN·m (10^5 dyn·cm), but brief intervals of 100 mN·m (torque booster) can be used for conditioning specimens. The standard temperature range is 5–95°C, but extension to lower and higher temperatures is possible. The Bohlin CSM is a controlled stress rheometer for polymer melts that can also operate in the oscillatory mode (10^{-3} to 30 Hz). The CSM features electrically heated plates with a temperature range of ambient to 400°C. The torque range is 10^{-3} to 20 mN·m. In addition to research viscometers, Bohlin also supplies a portable viscometer for laboratory and field use: the Visco 88 BV. This concentric cylinder instrument can measure viscosities of $6.5 - 3.5 \times 10^5$ mPa·s over a shear rate range of 4–1200 s^{-1} (which can be extended to 6000 s^{-1}) at temperatures of 0–100°C. The torque range is 5×10^{-2} to 10 mN·m.

C. W. Brabender Instruments, Inc. has long been a supplier of rheometers, mixers, extruders, and other instruments to many industries, including those manufacturing resins and plastics, rubber, food, building products, pigments, and adhesives. Brabender once supplied true Couette viscometers, the Rheotron and Viscotron, but in the 1990s have restricted themselves to production of instruments that combine rotation with mixing and/or extrusion. The Plasti-Corder PL 2000 and Electronic Plasti-Corder torque rheometers are quite different from most of the other commercial rotational viscometers and may fit better with the extrusion rheometers in Table 6. However, many of their applications are to mixing rather than extrusion and their use certainly is not restricted to polymer melts. These rheometers detect torque with a dynamometer consisting of a movable gear box coupled to a load cell by a torque arm. The measuring attachment is either a small batch mixer or a miniature extruder. Torque is plotted against working (mixing or extruding) time.

The Brookfield viscometer (Fig. 30), specified or mentioned in 20 different ASTM methods, is used for moderately low shear rate measurements. It is inexpensive, rugged, and easily used to characterize a wide range of materials. The original and most common sensor is a disk rotating in an infinite sea of fluid. A concentric cylinder system, eg, a Small Sample adapter, UL adapter, or Thermosel system, provides better-defined geometry and, in the case of the Small Sample adapter, good temperature control. There is also a cone-plate version, the Wells-Brookfield viscometer. With its well-defined geometry, good temperature control, and ease of cleaning, the cone-plate instrument is useful for characterizing paints, pigment pastes, inks, adhesives, and many other materials. The cone-plate and concentric cylinder systems are preferred for accurate measurements. The newer digital models with computer recorder-printer outputs are even easier to use than the old dial models and extend application to the characterization of time-dependent behavior. Combined with a PC and Brookfield's proprietary software, the DV-II+ and DV-III viscometers are powerful yet inexpensive rheological tools.

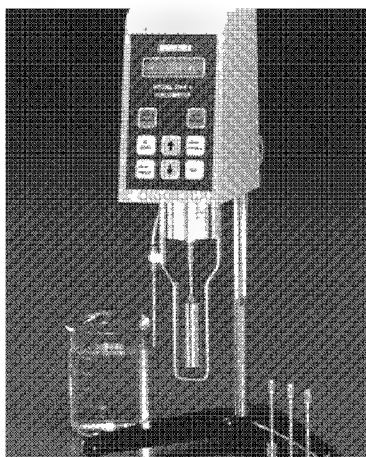


Fig. 30. Brookfield digital viscometer.

Courtesy of Brookfield Engineering Laboratories, Inc.

With several springs, which function as torque gauges, and a number of spindles, viscosities can be measured up to 10^9 mPa·s with the Brookfield viscometer. The shear rates depend on the model and the sensor system; they are ca. 0.1–100 s^{-1} for the disk spindles, $< 132 s^{-1}$ for concentric cylinders, and $< 1500 s^{-1}$ for the cone-plate for low viscosity samples. Viscosities at very low (ca. $10^{-3} - 1 s^{-1}$) shear rates can be measured with the concentric cylinder or the cone-plate system by a relaxation technique (158,177–179).

Brookfield has introduced a new digital cone-plate viscometer in two versions. The CAP 1000 is a single speed instrument (12,000 or 3,000 s^{-1} with 60 Hz current) that upgrades the ICI cone-plate design (ASTM D4287). The CAP 2000 is a multispeed viscometer with a viscosity range of 1–15,000 mPa·s. This instrument covers a wide range of shear rates (166–26,600 s^{-1}) and complements the low shear Wells-Brookfield viscometer. These two instruments form a relatively inexpensive package that allows the characterization of a large number of materials over a wide range of viscosities and shear rates. Brookfield has also developed a digital Stormer-type viscometer (ASTM D562), Model KU-1, which is an improvement over the old manual Stormer. This low shear ($\sim 50 s^{-1}$) viscometer is commonly used to test house paints.

The Cannon Instrument Company produces a line of rotational viscometers, most of which are quite specialized, eg, Cold Cranking Simulators (ASTM D5293) and Mini-Rotary viscometers (ASTM D3829 and D4684) for automotive engine oils. They also have a general use instrument similar to Brookfield's basic viscometer.

The German firm Coesfeld sells a range of viscometers (Visco-Mix and Rheosyst) designed primarily for the rubber and plastics industries. These instruments allow the measurement of viscosity during stirring or mixing without the need for removing individual specimens. Therefore, the rheological behavior of a fluid can be studied under simulated plant process conditions. It is also possible to use the viscometers for direct measurements in plant process streams. The viscometers can be computer-controlled through the addition of the Coesfeld Rheocalculator and associated software. A number of viscosity sensors are available, including concentric cylinders and anchor, propeller, and turbulent (Cowles-type) paddles.

The Ferranti-Shirley viscometer was the first commercial general-purpose cone-plate viscometer; many of the instruments still remain in use in the 1990s. Viscosities of 20 to 3×10^7 mPa·s can be measured over a shear rate range of 1.8–18,000 s^{-1} and at up to 200°C with special ceramic cones. Its features include accurate temperature measurement and good temperature control (thermocouples are embedded in the water-jacketed plate), electrical sensing of cone-plate contact, and a means of adjusting and locking the position of the cone and the plate in such a way that these two just touch. Many of the instruments have been interfaced with computers or microprocessors.

The Haake Rotovisco system has included a number of rotational viscometers that share the same wide range of sensors, both concentric cylinders and cone-plate. The standard research instrument, which took the place of several earlier viscometers, has been the RV 20, a versatile viscometer having a wide range of speeds and geometries. A number of packages are available, ranging from a basic (manual) system to a fully automatic, computerized system. Besides the standard concentric cylinder geometry, a cone-plate-parallel plate attachment (PK100) is available, as well as a low shear Couette device (CV100) and another cup-and-bob sensor (M5-Osc) for oscillatory measurements. Both the CV100 and PK100 can also carry out oscillatory measurements. The latter has several excellent features, including a built-in temperature probe on the surface of the measuring plate and an electronic micrometer with a digital display, which allows precise gap setting for parallel plate measurements. Special sensors are available for the RV 20 for high temperatures (up to 500°C) and high pressures (up to 100 MPa or 1000 atm) or a vacuum. Depending on the sensor, the RV 20 has viscosity and shear rate ranges of 1– 10^9 mPa·s and 4×10^{-4} to $4 \times 10^4 s^{-1}$, respectively.

Haake has introduced other viscometers, including the RheoStress RS100, which offers controlled stress as well as controlled shear rate and oscillatory modes over a temperature range of 50 to 350°C (ambient to 500°C is also possible). This versatile viscometer covers a shear rate range of 10^{-6}

to 10^3 s^{-1} , a torque range of 10^{-3} to $10 \text{ mN}\cdot\text{m}$, and a frequency (in the oscillatory mode) of 10^{-3} to 20 Hz . Haake also produces an inexpensive digital viscometer for quality control and routine measurements, the Viscotester VT50. This instrument has a viscosity range of $0.5\text{--}10^7 \text{ mPa}\cdot\text{s}$, a shear rate range of $0.1\text{--}10^4 \text{ s}^{-1}$, and operating temperatures of -50 to 200°C (external temperature control). A variety of sensors are available, although the concentric cylinder and cylinder in a sea of fluid are standard. The VT50 can be connected to a recorder or an IBM-compatible personal computer.

The Hercules viscometer was originally designed for paper and paperboard coatings, but its use has been extended to paints, adhesives, mineral slurries, emulsions, and starch solutions. The instrument, noted for being robust and reliable, is particularly well suited for quality control and product formulation. It is capable of measuring viscosity over a moderate range ($1\text{--}10^6 \text{ mPa}\cdot\text{s}$) up to high shear rates ($115,000 \text{ s}^{-1}$). A more recent model is the electronic ET 24-6, which has higher accuracy and better speed control than the older models; however, it lacks temperature control.

The ICI cone–plate device is a simple, inexpensive, one-speed instrument designed for routine determination of paint viscosities at a shear rate comparable to that of brushing or spraying ($12,000 \text{ s}^{-1}$ at 60 Hz and $10,000 \text{ s}^{-1}$ at 50 Hz) (180,181). The viscosity range at that shear rate is limited to $5\text{--}1000 \text{ mPa}\cdot\text{s}$. A wider-angle cone allows measurement to $10,000 \text{ mPa}\cdot\text{s}$, but at a lower shear rate of $3,000 \text{ s}^{-1}$. A lower shear rate version (600 s^{-1}) measures up to $5 \times 10^4 \text{ mPa}\cdot\text{s}$. Both instruments can be obtained with a temperature capability of $25\text{--}200^\circ\text{C}$ in 25°C steps, suitable for measuring the melt viscosities of some polymers. Brookfield in the 1990s offers similar viscometers as does Sheen Instruments Ltd. A low shear relaxation viscometer based on the ICI cone–plate has been reported (182) as well as a low frequency oscillating version (83). Neither is commercially available as of this writing (1996).

The MacMichael viscometer is probably the most straightforward rotational viscometer. The outer cup rotates and the inner cylinder is suspended from a torsion wire. The drag on the inner cylinder is measured as degree of twist on the wire. Wires of different stiffness are available, and the maximum viscosity is ca $10 \text{ mPa}\cdot\text{s}$. The shear rate range is limited, ca $2\text{--}12 \text{ s}^{-1}$, but with modification, higher shear rates can be attained. The instrument is best suited for slow, time-dependent effects and low shear viscosities.

In the Mooney shearing disk viscometer, a serrated disk is rotated in a sample fixed in a pressurized cavity. The instrument was developed for rubber and other elastomeric materials and is a standard quality control instrument in the rubber industry (ASTM D1646). It is used to measure high viscosities given in arbitrary Mooney units, but usually ca $7.5 \times 10^7 \text{ mPa}\cdot\text{s}$ at low (ca 1.5 s^{-1}) shear rates.

The Nametre Rotary B rotational viscometer measures torque in terms of the current needed to drive the d-c motor at a given speed while a material is under test. The standard sensors are coaxial cylinders or Brookfield disk-type spindles, but a cone–plate system is also available. The viscosity range for the coaxial cylinder sensors is 5 to $5 \times 10^4 \text{ mPa}\cdot\text{s}$, and the maximum shear rate is 200 s^{-1} .

The Physica rotational viscometer line includes a number of units, ranging from a simple, inexpensive manual viscometer (Rheolab MC1) to a highly sophisticated instrument (Rheolab MC200) capable of controlled stress and creep measurements as well as oscillation and viscosity–shear rate determinations. Good instruments for general use are the Viscolab LC3 and LC10 viscometers and the Rheolab MC10. The first two have viscosity ranges of 1 to $3 \times 10^8 \text{ mPa}\cdot\text{s}$ at shear rates of 0.1 to $5 \times 10^4 \text{ s}^{-1}$. The MC10 has a viscosity range of 1 to $3 \times 10^6 \text{ mPa}\cdot\text{s}$, and a shear rate range of $10^{-3}\text{--}15,000 \text{ s}^{-1}$. The Viscolab instruments have a temperature range of -40 to 180°C , but the MC10 can go to 500°C . The Physica Rheolab MC120 viscometer has wider viscosity and shear rate ranges (1 to $10^{12} \text{ mPa}\cdot\text{s}$ and 10^{-5} to $5 \times 10^4 \text{ s}^{-1}$, respectively) and cone–plate and parallel plate geometries as well as concentric cylinders. It can also carry out controlled stress and creep recovery measurements over a shear stress range of 7×10^{-2} to $3.5 \times 10^4 \text{ Pa}$ (torque range of $0\text{--}50 \text{ mN}\cdot\text{N}$). There is also a low stress LS 100 instrument that does oscillatory measurements as well as controlled stress.

The Ravenfield model BS viscometer is a wide shear rate range instrument with several possible measurement systems: cone–plate, parallel plates, concentric cylinders, and taper plug. The last gives shear rates of up to 10^6 s^{-1} , and the cone–plate of up to $8 \times 10^4 \text{ s}^{-1}$. The viscosity range is $10^2\text{--}10^8 \text{ mPa}\cdot\text{s}$. Measurements can normally be made up to 170°C , but with special modifications even higher temperatures can be achieved. A computer interface permits two-way communication with a computer.

The Rheometric Scientific family of complex dynamic viscometers incorporates microprocessors that provide test control and data analysis. The viscometer of choice for low to medium viscosity fluids is the widely used fluids spectrometer RFS II. This instrument is capable of measuring viscosity and elasticity (normal force, dynamic modulus) over a wide range of steady-state and dynamic shear. Cone–plate, parallel plate, and coaxial cylinder geometries can be used. Shear rates at steady shear are $10^{-1}\text{--}10^4 \text{ s}^{-1}$, and oscillatory shear frequencies are $10^{-2}\text{--}500 \text{ rad s}^{-1}$. Viscosities of $10\text{--}10^6 \text{ mPa}\cdot\text{s}$ can be measured. Rheometrics also markets a controlled stress rheometer (RSR) that is useful for determining yield stresses and characterizing creep behavior in polymer melts, adhesives, coatings, and viscous suspensions. The torque range is 5×10^{-3} to $20\text{--}50 \text{ mN}\cdot\text{m}$, and the standard test fixtures are 25-mm parallel plates, which give a shear stress range of $116\text{--}3200 \text{ Pa}$ at temperatures up to 350°C . Rheometric Scientific introduced a new Advanced Rheometric Expansion System (ARES) in the 1990s which allows custom configuration from a variety of actuators, transducers, fixtures, and environmental systems.

The Rheo-Tech International (RTI) viscoelastic rheometer is a controlled stress rheometer that measures strain and strain rates using steady, dynamic, and oscillatory modes to characterize rheological properties. It has a torque range of $10^{-3}\text{--}10 \text{ mNm}$, a viscosity range of 0.5 to $5 \times 10^8 \text{ mPa}\cdot\text{s}$, and a shear rate range of $10^{-6}\text{--}10^4 \text{ s}^{-1}$. A wide range of sensor geometries is available, and special systems can be designed to customer specifications. The Rheo-Tech VER can be connected to any computer through a suitable interface. Comprehensive software is available which is designed to take full control of the measuring head and to record all data.

The Stormer viscometer is a useful low shear rotational viscometer that can be converted to high shear with a suitable attachment. The original, manual version is one of the few simple instruments in which the shearing stress rather than the rate of shear is kept constant. A constant torque is applied to an inner cylinder or paddle, and the rate of rotation is measured. In the older models, the shear stress is applied by attaching weights to a string and permitting free fall through a vertical distance. Newer models (Thomas-Stormer ETS-1000 and Brookfield KU-1) are shear rate-driven rather than stress-driven. The instruments accept many different types of rotors and cups, including some from other viscometers, such as the MacMichael. Therefore, many different geometries are possible, including a narrow gap ($R_i/R_o = 0.99$) concentric cylinder geometry for high shear: the CRGI Glidden Miniviscometer (183,184).

The TA Instruments CSL² rheometer is a computer-driven controlled stress instrument formerly known as the Carri-Med CS rheometer. It is a versatile viscometer having the ability to make flow curve measurements as a function of shear rate, as well as shear stress, creep, and recovery measurements, and to simulate stress controlled processes. The viscosity range is $10^2\text{--}10^9 \text{ mPa}\cdot\text{s}$, the shear rate range 10 to $5 \times 10^3 \text{ s}^{-1}$, and the torque range $1 \mu\text{N}\cdot\text{m}$ to $50 \text{ mN}\cdot\text{m}$. The stress range is 10^{-2} to $2 \times 10^4 \text{ Pa}$. The low shear rate mode requires a series of creep measurements. This and the normal creep mode are useful for determining the structure and elasticity of industrial products such as paints, inks, and adhesives. The instrument also has a controlled stress oscillatory mode and can determine viscosity and elasticity over a frequency range of $10^{-4}\text{--}40 \text{ Hz}$. Cone–plate, parallel plate, concentric cylinder, and cone-in-cone geometries are available. A Peltier system provides excellent temperature control from -20 to 100°C for cone–plate and parallel plates. A circulating fluid system for concentric cylinders is available, typically from -30 to 150°C . Higher temperatures for polymer melts (up to 400°C) are possible with the Extended Temperature Module. This instrument has a number of useful features that are improvements over early models, including automatic cone–plate and parallel plate gap setting and the ability to make inertia corrections with low viscosity fluids.

The Weissenberg Rheogoniometer (49) is a complex dynamic viscometer that can measure elastic behavior as well as viscosity. It was the first rheometer designed to measure both shear and normal stresses and can be used for complete characterization of viscoelastic materials. Its capabilities include measurement of steady-state rotational shear within a viscosity range of $10^{-1}\text{--}10^3 \text{ mPa}\cdot\text{s}$ at shear rates of $10^{-4}\text{--}10^4 \text{ s}^{-1}$, of normal forces (elastic effect) exhibited by the material being sheared, and of an oscillatory shear range of 5×10^{-6} to 50 Hz , from which the elastic modulus and dynamic viscosity can be determined. A unique feature is its ability to superimpose oscillation on steady shear to provide dynamic measurements under flow conditions; all measurements can be made over a wide range of temperatures (-50 to 400°C).

The current family of Weissenberg Rheogoniometers differs from its predecessors in its use of solid-state electronics, digital displays, and computer interfacing. The four basic models are available as a manually operated system or an automated system controlled by a computer. A wide range of sensor geometries is available, including serrated plates and Mooney cylinders, as well as the more common geometries.

Moving Body Viscometers. In moving body viscometers, the motion of a ball, bubble, plate, needle, or rod through a material is monitored.

Falling ball viscometers are based on Stokes' law, which relates the viscosity of a Newtonian fluid to the velocity of the falling sphere. If a sphere is allowed to fall freely through a fluid, it accelerates until the viscous force is exactly the same as the gravitational force. The Stokes' equation relating viscosity to the fall of a solid body through a liquid may be written as equation 34, where r is the radius of the sphere; d_s and d_l are the density of the sphere and the liquid, respectively; g is the gravitational force; and v is the velocity of the sphere.

$$\eta = \frac{2r^2g(d_s - d_l)}{9v} \quad (34)$$

Viscometers of the falling ball type can be used over an extremely wide viscosity range, but are usually employed for fairly viscous materials because small balls and small differences in density are needed to obtain a suitably slow rate of fall in a low viscosity fluid. The devices are limited to measurements of Newtonian fluids because no practical formula has been developed for non-Newtonian materials. They had been considered as instruments for routine viscosity measurements rather than highly accurate work, but more recent designs (185–187) have changed this. The technique has proved to be useful in the study of suspensions, including those that are opaque and concentrated. The cylinder is jacketed for temperature control to within 0.1°C. Optical techniques are used for clear solutions and x-rays for opaque suspensions. A high speed video system is used for recording data.

The speed at which a sphere rolls down a cylindrical tube filled with a fluid or down an angled plate covered with a film of the fluid also gives a measure of viscosity. For the cylindrical tube geometry, equation 35, a generalized form of the Stokes' equation is used for any given instrument, where v is the translational velocity of the rolling sphere and k is the instrument constant determined by calibration with standard fluids.

$$\eta = \frac{k(d_s - d_l)}{v} \quad (35)$$

The Hoesppler viscometer is probably the most widely used falling-rolling sphere viscometer. It is composed of a water-jacketed, precision bore glass tube mounted at an angle 10° from the vertical. A series of balls of different diameters allow measurement of viscosities at 0.5–10⁶ mPa·s. A simple falling ball viscometer may be constructed from a test tube or a graduated cylinder with a rubber stopper and a short guide tube for insertion of the balls, as shown in Figure 31. Miniature ball bearings are suitable spheres. They may be obtained in different sizes and in materials of various densities.

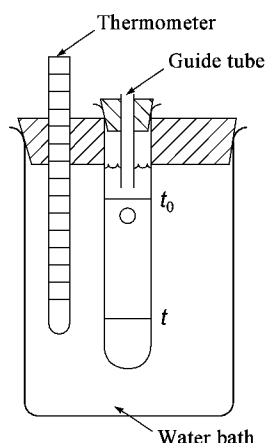


Fig. 31. Falling ball viscometer. Lines t_0 and t are the timing lines for velocity determinations.

More recent developments in the rolling ball area include an automated microviscometer, the Paar AMV 200, from Paar Physica. The specimen to be measured is introduced into a glass capillary down which a gold-covered steel ball rolls. The rolling time is measured automatically. The shear stress may be varied by changing the inclination angle of the capillary tube. The shear rate range is 10–1000 s⁻¹, which makes the instrument useful for non-Newtonian fluids, although it is limited to low viscosity materials (0.5–800 mm² s). The small specimen size (0.1–3 mL) allows testing of biological fluids and other materials that may be available only in small volumes. Paar Physica also has a magnetic field-driven rolling ball viscometer (MVM) that operates at much lower shear rates (down to 5 × 10⁻⁶ s⁻¹) and higher viscosities (10⁶ mPa·s).

The technique of determining viscosity from the velocity of a rolling ball on an inclined plate or panel has been used largely for following the flow, drying, and curing of paints and other coatings (188–190), but can also be used for other materials. The system can be calibrated with standard oils or other fluids of known viscosity. The geometry is ill defined because the ball often slides and rolls, but the technique is useful. Rolling ball methods based on an inclined plane or an inclined path leading to a horizontal test specimen have been employed to determine the tack of pressure-sensitive adhesives (ASTM D3121) (191–193). The reciprocal of the distance rolled by the ball is considered to be proportional to the tackiness of the adhesive.

In the Irvine-Park falling needle viscometer (FNV) (194), the moving body is a needle. A small-diameter glass or stainless steel needle falls vertically in a fluid. The viscous properties and density of the fluid are derived from the velocity of the needle. The technique is simple and useful for measuring low (down to 10⁻⁴ s⁻¹) shear viscosities. The FNV-100 is a manual instrument designed for the measurement of transparent Newtonian and non-Newtonian fluids. The FNV-200 is an automatic viscometer used for both transparent and opaque fluids. The latter has a magnetic sensor and uses needles containing small magnets. The falling needle viscometer has a viscosity range of 1 to 2.4 × 10⁶ mPa·s and a temperature range of -30 to 125°C.

Viscosity can also be determined from the rising rate of an air bubble through a liquid. This simple technique is widely used for routine viscosity measurements of Newtonian fluids. A bubble tube viscometer consists of a glass tube of a certain size to which liquid is added until a small air space remains at the top. The tube is then capped. When it is inverted, the air bubble rises through the liquid. The rise time in seconds may be taken as a measure of viscosity, or an approximate viscosity in mm² s may be calculated from it. In an older method that is commonly used, the rate of rise is matched to that of a member of a series of standards, eg, with that of the Gardner-Holdt bubble tubes. Unfortunately, this technique employs a nonlinear scale of letter designations and may be difficult to interpret.

Under a broad definition of a moving body, certain other viscometers can be considered to be of this type, including the band, falling rod, and sliding plate viscometers. The band viscometer includes an arrangement of parallel plates, the basic geometry used to define shear viscosity. The band, which is a strip of Mylar film, is sandwiched between two fixed plates. The fluid is placed between the band and the plates, and the band is pulled through the fluid. For a given gap and force on the band, the speed of the band is a measure of viscosity.

The falling rod viscometer, though similar to the band viscometer, is based on the movement of a rod rather than a plate through the fluid (195–198). It is a form of a falling coaxial cylinder viscometer. Initially, this device was used for semiempirical studies of materials such as bitumens and rosins. In the 1990s, the Laray falling rod viscometer (Fig. 32) has become a standard test instrument in the ink industry (ASTM D4040), and more recent versions of the falling rod viscometer are capable of precise measurements of polymer melts and solutions (199).

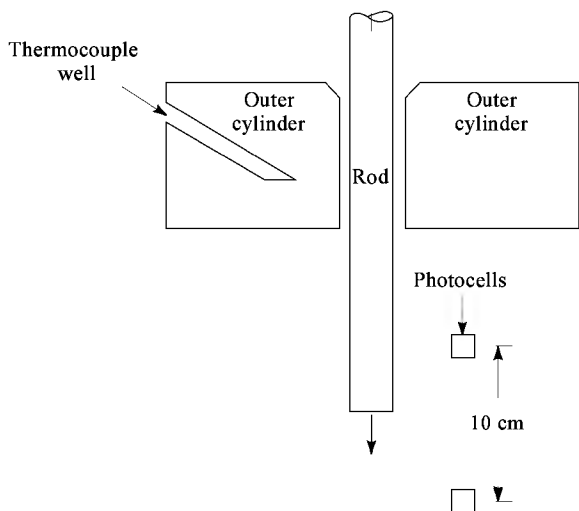


Fig. 32. Schematic of the principal parts of the Laray falling rod viscometer (197). The rod moves through the cylindrical hole in the outer cylinder. During measurements there is a thin layer of fluid in the gap between the rod and the cylinder. The photocells are used to time the motion of the rod.

Courtesy of the Oil and Colour Chemists' Association.

The sliding plate rheometer has been in use for some time, but new types have been developed in the 1990s (200–204). The well-defined simple shear deformation generated by this technique is useful for the characterization of polymer melts and concentrated solutions. Of particular interest is the fact that it can cause a high degree of molecular stretching. A commercial instrument, the Intedaken True Shear Rheometer, has come from this work. This rheometer can measure viscosity under simple, interrupted, exponential, or oscillatory shear. The shear rate range is 0.05 to 200 s⁻¹ and temperatures can go up to 300°C. The sliding plate rheometer is useful for the characterization of polymer slip at solid boundaries and the determination of the effectiveness of processing aids (205,206).

Other Viscometers. A number of other viscometers are built for specific research or product applications. In one design, vibrational techniques are used to measure viscosity. The Bendix Ultra-Viscoson, the Automation Products Dynatrol, and the Nametre are all based on this principle. The Ultra-Viscoson determines viscosity based on the time required for longitudinal vibrations of an immersed rod to subside. The Dynatrol determines viscosity by measuring the amplitude of vibration of an immersed flexural member. The Nametre viscometer measures viscosity by determining the power required to maintain torsional vibrations of a small stainless steel spherical cylinder or rod (207). It has a temperature range of -40 to 230°C. Because the rate of shear is not easily determined or changed, these instruments are best used for controlling or studying processes in which viscosity changes with time or temperature. In such cases they can be useful because wide ranges of viscosity can be measured without changing sensors (10⁻¹ to 10⁶ mPa·s for the Nametre). These devices are employed for in-line process control for many polymeric solution-based products because of their insensitivity to stirring and flow, wide viscosity ranges, simplicity of design, and ability to be controlled from and transmit data to a remote point.

Measurement Techniques

Extensional Viscosity. All three types of extensional viscosity can be measured (101,103): uniaxial, biaxial, and pure shear. Only a few commercial instruments are available, however, and most measurements are made with improvised equipment. Extensional viscosity of polymer melts can be estimated from converging flow (entrance pressure) or from a melt strength drawdown test (208).

Most uniaxial measurement techniques involve extending a strand or cylindrical rod of the material and measuring the force required. To measure fluids, the fluid is extruded from a spinnerette nozzle and allowed to fall under gravity. It is then extended by being rolled up on a rotating drum (101,109,209–212). The generated force is measured by the deflection of the nozzle or tube as the latter is pulled by the filament of fluid or at the takeup roller. This method was the basis for the Carri-Med spin line rheometer (SLR) (209,212) shown in Figure 33. The fluid is kept in a reservoir, which is pressurized by a clean gas. The pressure forces the fluid along a calibrated, thin-walled tube and out through a spinnerette nozzle. The fluid falls vertically, is taken up tangentially on a rotating drum, and is elongated. The rate of deformation of the specimen is determined from the filament deflection and width during spinning. Controllable variables are the nozzle diameter, flow rate, drum speed, and length of filament. The whole assembly is housed in an environmental chamber, and the specimen can be heated to 120°C. Low viscosity fluids require a takeup device, a vacuum attachment which sucks up the liquid. This reduces the lower viscosity limit for the instrument from 500 to about 10 mPa·s

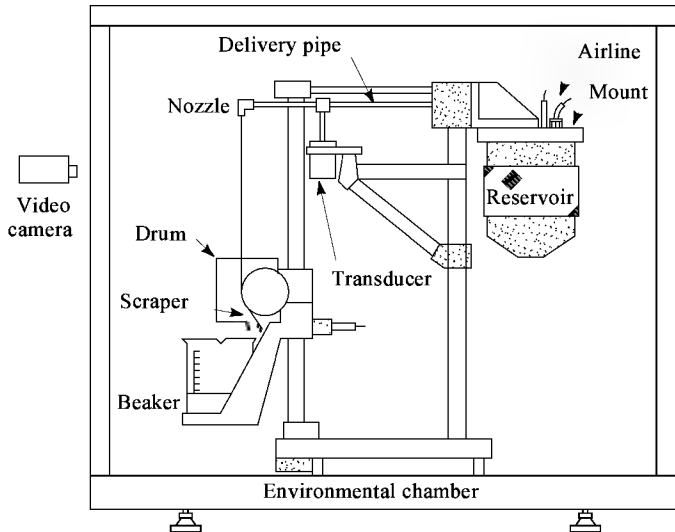


Fig. 33. Diagram of spin line rheometer (212).

Courtesy of Rheologica Acta.

Another type of experiment involves a fluid filament being drawn upward against gravity from a reservoir of the fluid (101,213,214), a phenomenon often called the tubeless siphon. The maximum height of the siphon is a measure of the spinnability and extensional viscosity of the fluid. More quantitative measures of stress, strain, and strain rate can be determined from the pressure difference and filament diameter. A more recent filament stretching device in which the specimen is held between two disks that move apart allows measurements in low viscosity liquids (215). All of these methods are limited to spinnable fluids under small total strains and strain rates. High strain rates tend to break the column or filament.

There is another technique that can be used with low viscosity (100 mPa·s) nonspinnable fluids and which allows high strain rates ($>10^3 \text{ s}^{-1}$). A modified opposing jets device (216) is employed that consists of opposing nozzles through which the fluid is sucked or blown out. Extensional viscosity is determined from the force required to keep the nozzles at a fixed distance apart as a function of flow rate. A wide range of fluids can be investigated, and the high strain rates obtainable make it possible to study industrial processes and the effects of low concentrations of additives. A commercial instrument (Rheometrics RFX) based on this device is on the market and has been used for some interesting research (217).

A sliding plate rheometer (simple shear) can be used to study the response of polymeric liquids to extension-like deformations involving larger strains and strain rates than can be employed in most uniaxial extensional measurements (56,200–204). The technique requires knowledge of both shear stress and the first normal stress difference, $N_1(\dot{\gamma})$, but has considerable potential for characterizing extensional behavior under conditions closely related to those in industrial processes.

A method for measuring the uniaxial extensional viscosity of polymer solids and melts uses a tensile tester in a liquid oil bath to remove effects of gravity and provide temperature control; cylindrical rods are used as specimens (218,219). The rod extruder may be part of the apparatus and may be combined with a device for clamping the extruded material (220). However, most of the more recent versions use prepared rods, which are placed in the apparatus and heated to soften or melt the polymer (103,111,221–223). A constant stress or a constant strain rate is applied, and the resultant extensional strain rate or stress, respectively, is measured. Similar techniques are used to study biaxial extension (101).

In another elongational rheometer (221), the specimen is suspended on the end of a flexible tape, which is wound onto a wheel turned by a servo-controlled torque motor. This design is the basis for the Gottfert Rheostrain instrument, with which strain rates of 5×10^{-3} to 2 s^{-1} are possible.

The Rheometric Melt Elongational Rheometer (RME) (Fig. 34) is designed for high elongations of polymer melts. It uses a small ($\sim 1 \text{ g}$) compression-molded specimen which it supports on a stream of inert gas and holds in place with a rotary clamping device. The specimen is heated to the test temperature (up to 350°C) by electric heaters and extended by the motion of metal belts. Three sets of belts are available to provide different ranges of extensional strain rates: 10^{-3} to 10^{-1} s^{-1} , 3×10^{-3} to $3 \times 10^{-1} \text{ s}^{-1}$, and 10^{-2} to 1 s^{-1} . The force generated by the specimen is measured by a spring-type transducer having a range of 10^{-3} to 2 N .

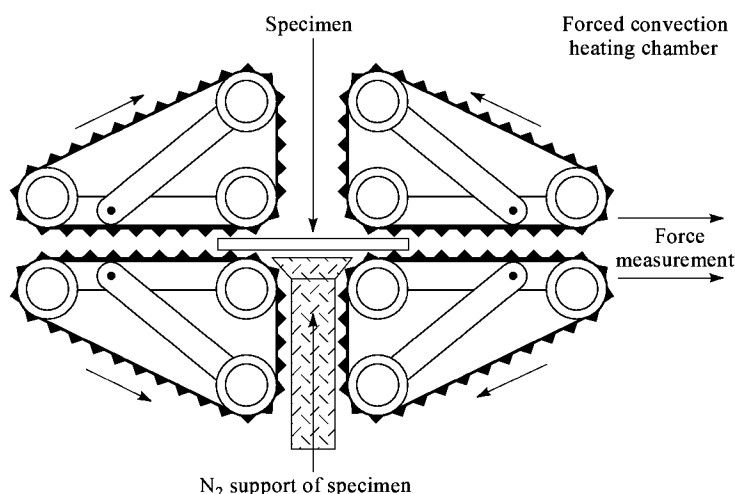


Fig. 34. Diagram of the Rheometric Melt Elongational Rheometer (RME) for elongational viscosity measurements on polymer melts.

Courtesy of Rheometric Scientific.

Biaxial extensional viscosity can be measured by several methods. The most common is bubble inflation, in which a thin polymer sheet is inflated with an inert gas or liquid (224–226). Another method involves stretching a polymer sheet using eight rotating clamps placed in an octagonal pattern (227). A complex servo control system is required to coordinate the motion of the clamps. In a newer technique, displacement is measured as the material is squeezed between two disks having lubricated surfaces (lubricated squeezing flow) (228–237). A commercial instrument based on this technique, the MARS-III Multifunction Axial Rheometer System, has been developed (232). The specimen is a disk 1–4 cm in diameter and 1–2 cm in thickness. A variety of deformation modes that emulate industrial processes are provided: deformation under constant load, constant strain rate, or constant velocity, or in a step mode for stress relaxation. After deformation the specimen can be monitored for elastic recovery. The temperature ranges are 20 – 250°C in oil and 20 – 400°C in air. The viscosity range is 10^4 to $10^{10} \text{ Pa}\cdot\text{s}$, with strain rates of 2×10^{-5} to 20 s^{-1} . The instrument is computer controlled and data collection is as automatic as the operator wishes.

Extensional viscosity that results purely from shear deformation seems to be of less interest, but has been measured (108). The rheology of several different polymer melts in terms of shear viscosity and uniaxial and biaxial extensional viscosity has been compared (231). Additional information on the measurement of extensional viscosity are also available (105,238–240).

Viscoelastic Measurement. A number of methods measure the various quantities that describe viscoelastic behavior. Some require expensive commercial rheometers, others depend on custom-made research instruments, and a few require only simple devices. Even qualitative observations can be useful in the case of polymer melts, paints, and resins, where elasticity may indicate an inferior batch or unusable formulation. For example, the extrusion swell of a material from a syringe can be observed with a microscope. The Weissenberg effect is seen in the separation of a cone and plate during viscosity measurements or the climbing of a resin up the stirrer shaft during polymerization or mixing.

Creep. Creep experiments are among the simplest for describing viscoelastic behavior. They involve the measurement of deformation as a function of time after a given load has been applied. Such measurements may be made on specimens in tension, compression, or shear. A simple apparatus for tensile creep measurements is shown in Figure 35 (241). The specimen is held between two film clamps: the upper is attached to a support, and the lower bears a weight. The apparatus can easily be placed inside a test tube and immersed in a water bath for temperature control. The extension of the sample is measured as a function of time. A graduated scale can be used to measure large extensions; for small extensions a cathetometer is required. Many more sophisticated devices exist, including tensile testers. The data are usually presented in the form of a strain–time curve (Figs. 16 and 36). Most materials undergo initial elastic deformation followed by a slow increase in strain with time. After the stress is removed, elastic deformation is recovered, but viscous deformation is not. Creep measurements are useful for evaluating the behavior of structural materials, fasteners, gaskets, pipe, and other articles that are subject to loading while in use. Creep measurements can also be carried out on fluids and are useful for measuring low shear viscosity and elasticity and for estimating and comparing structures, particularly in dispersions such as paints and inks.

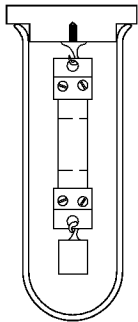


Fig. 35. Apparatus for creep measurements (241).

Courtesy of Marcel Dekker, Inc.

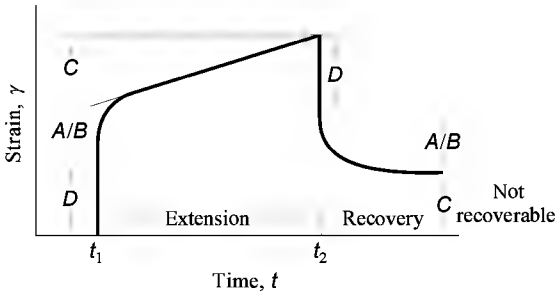


Fig. 36. Typical creep curve for a viscoelastic material. Stress applied at time t_1 and removed at t_2 .

Figure 36 is representative of creep and recovery curves for viscoelastic fluids. Such a curve is obtained when a stress is placed on the specimen and the deformation is monitored as a function of time. During the experiment the stress is removed, and the specimen, if it can, is free to recover. The slope of the linear portion of the creep curve gives the shear rate, and the viscosity is the applied stress divided by the slope. A steep slope indicates a low viscosity, and a gradual slope a high viscosity. The recovery part of Figure 36 shows that the specimen was viscoelastic because relaxation took place and some of the strain was recovered. A purely viscous material would not have shown any recovery, as shown in Figure 16b.

Stress Relaxation. In a stress-relaxation test the deformation is held constant, and the resulting stress is measured as a function of time. Deformation produces an initial stress that decays with time in the case of viscoelastic materials. A diagram of a simple stress-relaxation apparatus is shown in Figure 37. The sample is held by a set of clamps. The upper clamp is attached to a strain gauge or balance arm. The lower clamp is attached to a device for rapidly elongating the sample by a fixed amount. In an experiment the sample is rapidly deformed, and the resulting force measured as a function of time. The data may be plotted as stress, stress initial stress, or modulus. A stress-relaxation curve for a lightly cross-linked rubber is shown in Figure 38 (242).

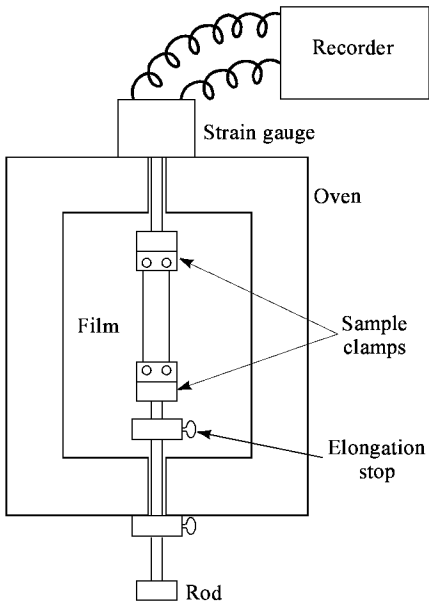


Fig. 37. Schematic diagram of a stress-relaxation apparatus (241).

Courtesy of Marcel Dekker, Inc.

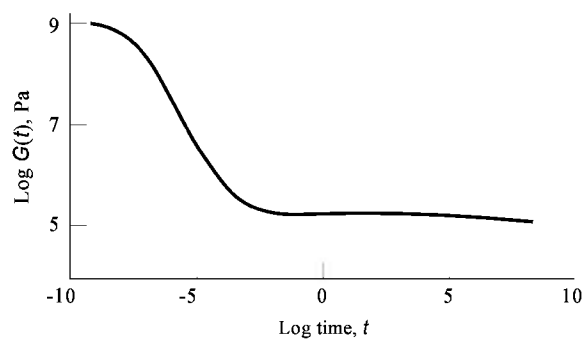


Fig. 38. Stress–relaxation curve for a lightly vulcanized rubber (242). To convert Pa to dyn cm², multiply by 10.

The stress–relaxation process is governed by a number of different molecular motions. To resolve them, the thermally stimulated creep (TSCr) method was developed, which consists of the following steps. (1) The specimen is subjected to a given stress at a temperature T for a time t , both chosen to allow complete orientation of the mobile units that one wishes to consider. (2) The temperature is then lowered to $T_0 \ll T$, where any molecular motion is completely hindered; then the stress is removed. (3) The specimen is subsequently heated at a controlled rate. The mobile units reorient according to the available relaxation modes. The strain, its time derivative, and the temperature are recorded versus time. By running a series of experiments at different orientation temperatures and plotting the time derivative of the strain rate observed on heating versus the temperature, various relaxational processes are revealed as peaks (243).

In a similar fashion, Thermally Stimulated Current spectrometry (TSC) makes use of an applied d-c potential that acts as the stress to orient dipoles. The temperature is then lowered to trap these dipoles, and small electrical currents are measured during heating as the dipoles relax. The resulting relaxation maps have been related to G' and G'' curves obtained by dynamic mechanical analysis (244–246). This technique, long carried out only in laboratory-built instruments, is available as a commercial TSC spectrometer from Thermold Partners L.P., formerly Solomat Instruments (247).

Penetration–Indentation. Penetration and indentation tests have long been used to characterize viscoelastic materials such as asphalt, rubber, plastics, and coatings. The basic test consists of pressing an indenter of prescribed geometry against the test surface. Most instruments have an indenting tip, eg, cone, needle, or hemisphere, attached to a short rod that is held vertically. The load is controlled at some constant value, and the time of indentation is specified; the size or depth of the indentation is measured. Instruments have been built which allow loads as low as 10^{-5} N with penetration depths less than 10^{-4} mm. The entire experiment is carried out in the vacuum chamber of a scanning electron microscope with which the penetration is monitored (248).

Most instruments make use of a probe geometry which gives an increasing area of contact as penetration proceeds. In this way, at some depth of penetration, the resisting force can become sufficient to balance the applied force on the indenter. Unfortunately, many geometries, eg, diamonds, pyramids, and cones, do not permit the calculation of basic viscoelastic quantities from the results. Penetrometers of this type include the Pfund, Rockwell, Tukon, and Buchholz testers, used to measure indentation hardness which is dependent on modulus.

The Rheo-Tex rheometer is an inexpensive, automated instrument using load cell technology to measure indentation and creep. Available software calculates hardness softness, brittleness, plasticity, and tensile strength. This instrument is particularly valuable for measurements on foods and personal care products.

Other penetrometer–indentometers (Fig. 39) include transducers to sense the position and movement of the probe and microprocessors for temperature control and data collection and reduction. These instruments are used mainly to measure softening points, which are not glass transitions but are usually close to those values. Because a softening point is indicative of behavior under load, it is often more useful for predicting performance than the T_g . Penetrometer–indentometers can also be used to measure indentation hardness, creep, creep recovery, and modulus. Examples of such instruments include the TA Instruments, Mettler, Perkin-Elmer, Seiko, and Shimadzu thermomechanical analyzers (TMAs), and the ICI Microindentometer. They can be used to generate modulus and modulus–temperature data from indentation–time plots by applying the Hertz equation (eq. 36) (144,249), where E is the elastic or Young's modulus, μ the Poisson ratio, r the radius of the hemispherical indenter, P the force on the indenter (mass load $\times g$), b the indentation, and H_k the indentation hardness.

$$\frac{E}{1-\mu^2} = \frac{3}{4r^{1/2}} \frac{P}{h^{3/2}} \quad H_k \tag{36}$$

Because the indentation varies with time, the modulus must be specified for a certain indentation time, eg, a 10-s modulus. The Hertz equation holds only for purely elastic materials. However, it has been applied to viscoelastic materials, including polymers and coatings, with excellent results (249–256). Indentation hardness vs temperature curves are shown in Figure 40 (249,251).

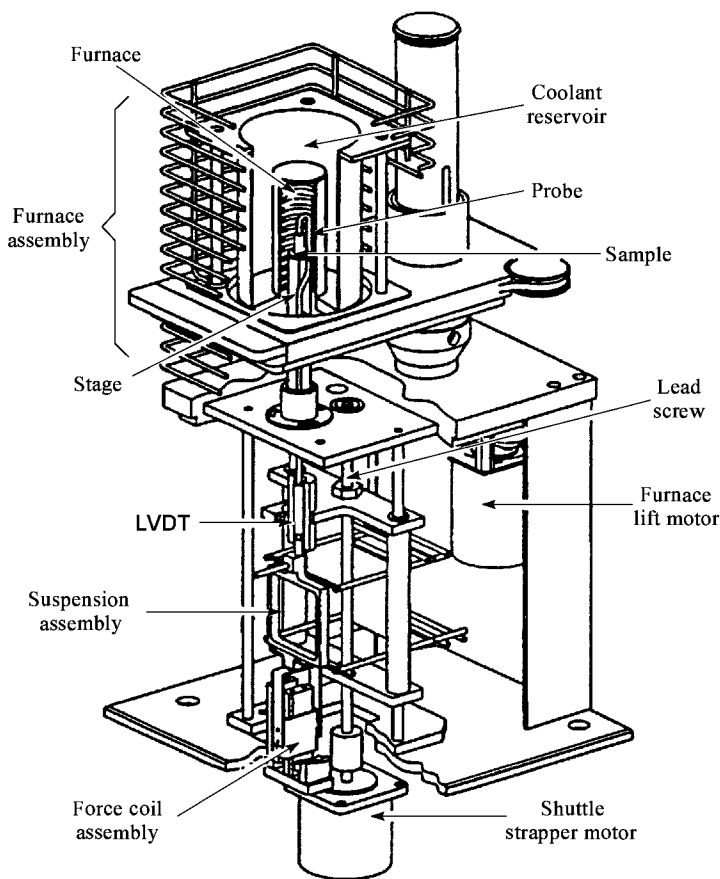


Fig. 39. Schematic of the TA Instruments model 2940 thermomechanical analyzer. LVDT = linear variable differential transducer.

Courtesy of TA Instruments, Inc.

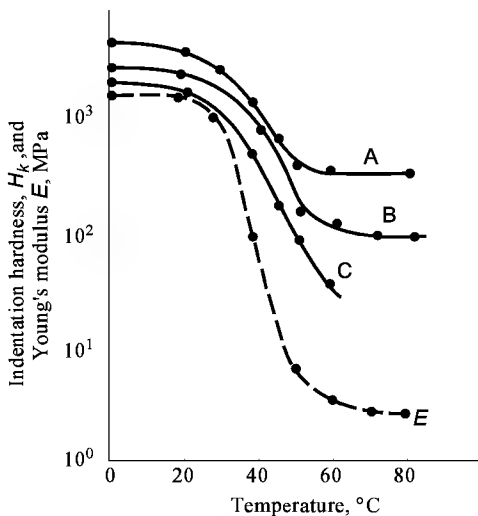


Fig. 40. Indentation hardness, H_k , at A, 19 μm ; B, 47 μm ; and C, 108 μm thickness, and Young's modulus, E , of a free film (as functions of temperature) of an acrylic coating (249). These results show the dependence of thickness that occurs with thin films. To convert MPa to psi, multiply by 145.

Courtesy of Farbe + Lack.

A fully automated microscale indenter known as the Nano Indenter is available from Nano Instruments (257–259). Used with the Berkovich diamond indenter, this system has load and displacement resolutions of 0.3 N and 0.16 nm, respectively. Multiple indentations can be made on one specimen with spatial accuracy of better than ± 200 nm using a computer controlled sample manipulation table. This allows spatial mapping of mechanical properties. Hardness and elastic modulus are typically measured (259,260) but time-dependent phenomena such as creep and adhesive strength can also be monitored.

Indentation is not usually thought of as a method for measuring viscoelastic properties, but the technique can be useful particularly for measurements of coatings on rigid substrates. Modern instrumentation, as described above, allows for precise, computer-controlled loading and unloading of force. The Mettler TMA, Perkin-Elmer's DMA, and the Nano Indenter have modes by which periodic and very short indentation creep measurements can be made while the specimen is being heated, thereby allowing estimates of the changes in modulus with temperature. A more quantitative technique is called modulus profiling, which uses a modified TMA instrument to measure tensile compliance, a measurement closely related to the inverse of the tensile modulus (261). The instrument can be used to profile or map tensile properties of a polymer specimen to show spatial variations resulting from processing inhomogeneities or aging phenomena such as oxidation.

Tensile Testing. The most widely used instrument for measuring the viscoelastic properties of solids is the tensile tester or stress-strain instrument, which extends a sample at constant rate and records the stress. Creep and stress-relaxation can also be measured. Numerous commercial instruments of various sizes and capacities are available. They vary greatly in terms of automation, from manually operated to completely computer controlled. Some have temperature chambers, which allow measurements over a range of temperatures. Manufacturers include Instron, MTS, Tinius Olsen, Applied Test Systems, Thwing-Albert, Shimadzu, GRC Instruments, SATEC Systems, Inc., and Monsanto.

A typical stress-strain curve generated by a tensile tester is shown in Figure 41. Creep and stress-relaxation results are essentially the same as those described above. Regarding stress-strain diagrams and from the standpoint of measuring viscoelastic properties, the early part of the curve, ie, the region

of small deformations, is usually of the most interest. At small strains, stress is proportional to strain, therefore the modulus is constant. At higher stresses and corresponding strains, a point is reached where the modulus depends on strain. Eventually, irreversible changes take place, resulting in rupture. These forms of behavior are shown in Figure 41. The linear relationship between stress and strain is shown in the region between the start of the experiment and point *A*. The latter is the yield point common to many materials, in which the specimen yields with no increase or with a decrease in stress. At higher levels the sample breaks at break point *B*. The tensile strength *b* and elongation at break *d* are important mechanical properties, used to characterize plastics, paint films, and many other materials. For more information on tensile testing, consult ASTM methods D2370, D638M, and D882 as well as more recent literature (262–266)

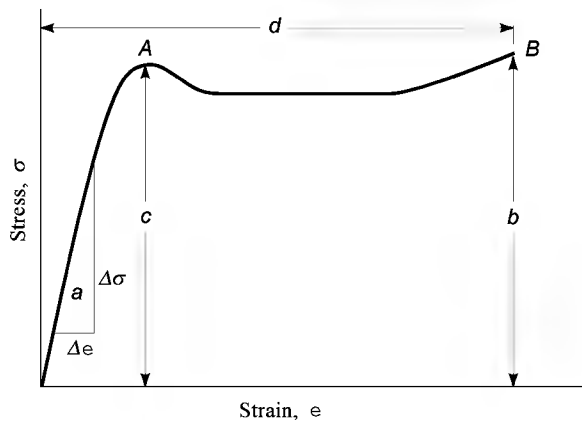


Fig. 41. Typical stress–strain curve. Point *A* is the yield point of the material; the sample breaks at point *B*. Mechanical properties are identified as follows: *a* $\Delta\sigma/\Delta\epsilon$, modulus; *b* tensile strength; *c* yield strength; *d* elongation at break. The toughness or work to break is the area under the curve.

Acoustic Measurements. Measurement of the propagation of ultrasonic acoustic waves has been found useful for determining the viscoelastic properties of thin films of adhesives. In this method, the specimen is clamped between transmitting and receiving transducers. The change in pulse shape between successive reverberation of the pulse is dependent on the viscoelastic properties of the transmitting material. Modulus values can be calculated (267,268).

Dynamic Measurements. Dynamic methods are required for investigating the response of a material to rapid processes, studying fluids, or examining a solid as it passes through a transition region. Such techniques impart cyclic motion to a specimen and measure the resultant response. Dynamic techniques are used to determine storage and loss moduli, G' and G'' , respectively, and the loss tangent, $\tan \delta$. Some instruments are sensitive enough for the study of liquids and can be used to measure the dynamic viscosity η' . Measurements are made as a function of temperature, time, or frequency, and results can be used to determine transitions and chemical reactions as well as the properties noted above. Dynamic mechanical techniques for solids can be grouped into three main areas: free vibration, resonance-forced vibrations, and nonresonance-forced vibrations. Dynamic techniques have been described in detail (242,251,255,266,269–279). A number of instruments are listed in Table 8. Related ASTM standards are listed in Table 9.

Table 8. Selected Commercial Dynamic Viscoelastometers

Instrument	Price range ^a	Frequency range, Hz	Approximate modulus range, Pa	Manufacturer
Autovibron Rheovibron	C	3.5–110 (0.01 to 1.0 optional)	10^5 – 10^{11}	Imass, Inc., Accord (Hingham), Mass.
Bohlin VOR	C	10^{-4} –10		Bohlin Reologi A.B., Lund, Sweden; Cranbury, N.J.
Mechanical Spectrometer	C	10^{-4} – 10^2		Bohlin Reologi A.B.
Carri-Med CS	C	10^{-7} –10	8×10^{-3} – 2×10^4	TA Instruments, New Castle, Del.
Dynamic Mechanical Analyzer (DMA)	C	10^{-2} – 2×10^2	10^3 – 10^{12}	TA Instruments
Dynamic Mechanical Thermal Analyzer ^b (DMTA)	C	10^{-2} – 2×10^2	10^3 – 10^{11}	Rheometric Scientific, Inc., Piscataway, N.J.
Dynastat Mark II Dynalyzer	C	10^{-2} – 10^2	10^4 – 10^{11}	Imass, Inc.
Mechanical Energy Resolver	C	0.5–60		Imass, Inc.
Metravib Micromecanalyzer	D	10^{-5} –1		Metravib Instruments, Dardilly, France
MVA Mach I (Viscoelasticimeter)	C	7– 10^3	10^3 – 10^{12}	Metravib
MVA Mach II (Viscoanalyser)	D	5– 10^3	10^3 – 10^{12}	Metravib
Rank Pulse Shearometer	A	~250	50 – 5×10^6	Rank Bros., Ltd., Cambridge, U.K. ^c
RDA II Dynamic Analyzer	C	10^{-5} –500 ^d	10^2 –10	Rheometric Scientific
RFS II Fluids Spectrometer	C-D	10^{-5} –100 ^d		Rheometric Scientific
RMS-800 Mechanical Spectrometer	D	10^{-3} –100 ^d	8×10^3 – 100^7	Rheometric Scientific
RSA II Solids Analyzer	C	10^{-3} –100 ^d		Rheometric Scientific
Rheo-Tech Viscoelastic Rheometer	B-C	10^{-1} – 10^2 ^d		Rheo-Tech International Ltd., London, U.K.
Dynamic Mechanical Analyzer DMA7	C	10^{-2} –50		Perkin-Elmer Corp., Norwalk, Conn.
DMS 110 120 Bending Shear Module	C	10^{-2} – 10^2	10^5 – 10^{12}	Seiko Instruments, Horsham, Pa.
DMS 200 210 Tension–Compression Module	C	10^{-2} – 10^2	10^5 – 10^{12}	Seiko Instruments
Torsional Braid Analyzer (TBA)	B	10^{-1} –1		Plastics Analysis Instruments, Princeton, N.J.
Weissenberg Rheogoniometer	D	5×10^{-6} –670	10^{-5} – 10^7	TA Instruments

^a 1995 Price ranges: A = \$10,000–\$25,000; B = \$25,000–\$50,000; C = \$50,000–\$100,000; D = > \$100,000 .

^b Formerly Polymer Laboratories, Ltd.

^c U.S. representative: Pen-Kem, Inc., Bedford Hill, N.Y.

^d Frequency range = rad s.

Table 9. ASTM Standards on Dynamic Mechanical Analysis

ASTM designation	Title
D4440	Rheological Measurement of Polymer Melts Using Dynamic Mechanical Procedures
D4065	Determining and Reporting Dynamic Mechanical Properties of Plastics
D5023	Measuring the Dynamic Mechanical Properties of Plastics Using Three-Point Bending
D5026	Measuring the Dynamic Mechanical Properties of Plastics in Tension
D-4 Proposed P244	Rheological Measurements of Bitumens Using Dynamic Shear Rheometers

Free-Vibration Methods. Free-vibration instruments subject a specimen to a displacement and allow it to vibrate freely. The oscillations are monitored for frequency and damping characteristics as they disappear. The displacement is repeated again and again as the specimen is heated or cooled. The results are used to calculate storage and loss modulus data. The torsional pendulum and torsional braid analyzer (TBA) are examples of free-vibration instruments.

The torsional pendulum (Fig. 42) is a simple apparatus for making dynamic tests in the frequency range of 0.1–10 Hz. The specimen is rigidly clamped at its lower end, and the upper end is clamped to an inertia bar or disk that is free to rotate. The suspension wire supporting the system is passed over a pulley, and the weight is counterbalanced to prevent tensile stresses on the specimen. The experiment is begun by setting the inertia member into free oscillation, and the natural frequency and decay of these oscillations are measured. The storage modulus, G' , is calculated from the square of the frequency, and the loss tangent, $\tan \delta$, from the logarithmic decrement of the sample. The logarithmic decrement is the natural logarithm of the ratio of two successive amplitudes.

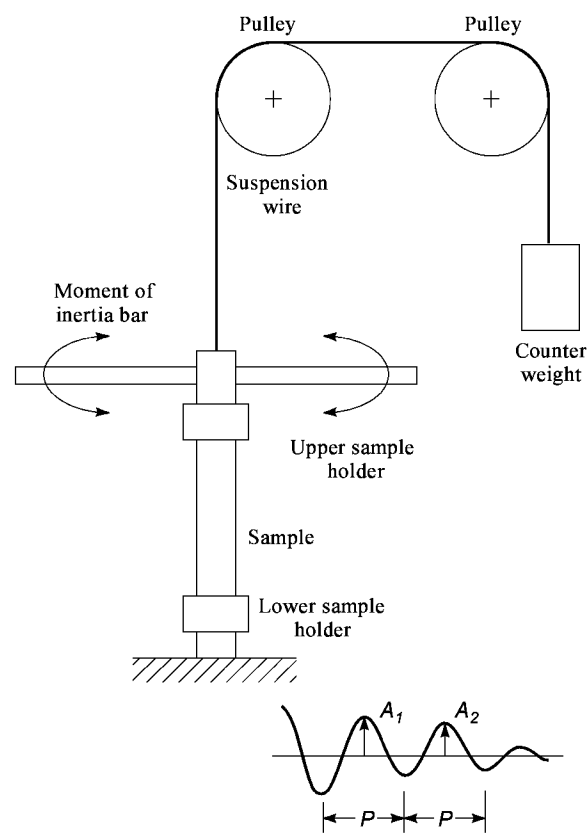


Fig. 42. Torsion pendulum and typical damped sine wave output. P is the period of the motion; A_1 and A_2 are successive amplitudes (241).

Courtesy of Marcel Dekker, Inc.

A Torsional Braid Analyzer (Fig. 43) is a torsional pendulum with which a composite specimen instead of a free film is used (269,273,274). The specimen is prepared by soaking a multifilamented glass braid in a polymer solution and, after mounting the specimen in the apparatus, removing the solvent by heating. Measurements are similar to those with the torsional pendulum. Rigidity and damping values are calculated. The former is proportional to G' ; the latter, to $\tan \delta$.

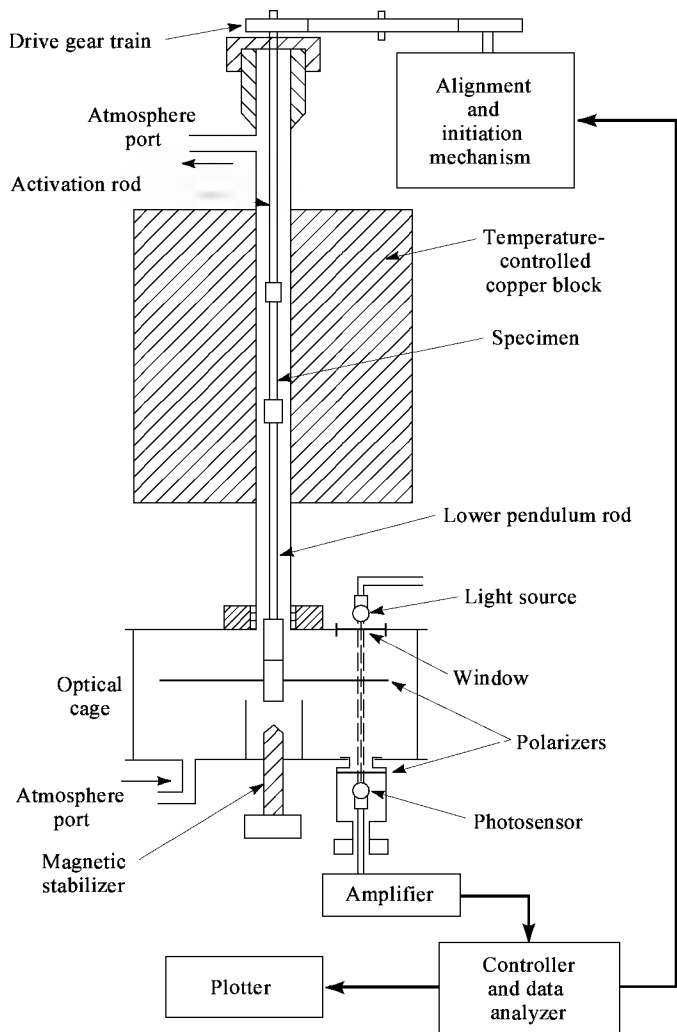


Fig. 43. Torsional braid analyzer (TBA).

Courtesy of Plastics Analysis Instruments, Inc.

With both instruments, data can be plotted as a function of time or temperature to give dynamic mechanical spectra. An example is shown in Figure 44, where rigidity and damping from TBA experiments are shown as functions of temperature for a styrene–butadiene–styrene radial block copolymer (280). Such spectra are particularly useful for the study of polymers, including the identification of physical transitions and the onset and kinetics of cross-linking (274,281) or embrittlement (280), as well as for measuring various viscoelastic parameters. The curves in Figure 44 show that the material has two glass-transition temperatures, each indicated by a downturn in the rigidity curve and a damping peak. The material is glassy (high modulus) below the lower T_g , rubbery or leathery (moderate modulus) between the two transitions, and almost fluid (low modulus) above the upper T_g . Knowledge of such characteristics is important for predicting behavior under processing or use conditions.

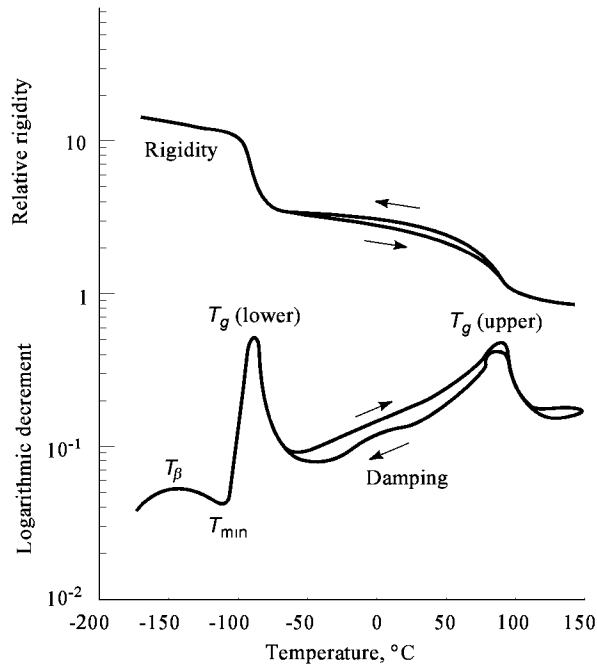


Fig. 44. Thermal mechanical behavior of a styrene–butadiene–styrene block copolymer in nitrogen at 180 to 150°C (280).

Courtesy of the Federation of Societies for Coatings Technology.

Resonance-Forced Vibration. Resonance-forced vibration devices drive the vibration of the specimen. This can be over a range of frequencies that includes the resonant frequency, which is detected as a maximum in the amplitude, or the instrument can be designed to detect the resonant frequency and drive the specimen at that frequency. An example of the resonance-forced vibration technique is the vibrating reed. A specimen in

the form of a thin strip, the reed, is clamped firmly at one end, and the other end remains free. The clamped end is forced to vibrate, and the amplitude of the oscillating free end is measured. As the driver frequency is varied, the natural resonance of the reed is reached, and the amplitude passes through a maximum. Young's modulus can be calculated from the square of the resonant frequency and the geometry of the specimen. The loss tangent can be calculated from the width of the resonance peak.

Another resonant frequency instrument is the TA Instruments dynamic mechanical analyzer (DMA). A bar-like specimen is clamped between two pivoted arms and sinusoidally oscillated at its resonant frequency with an amplitude selected by the operator. An amount of energy equal to that dissipated by the specimen is added on each cycle to maintain a constant amplitude. The flexural modulus, E' , is calculated from the resonant frequency, and the makeup energy represents a damping function, which can be related to the loss modulus, E'' . A newer version of this instrument, the TA Instruments 983 DMA, can also make measurements at fixed frequencies as well as creep and stress-relaxation measurements.

Nonresonance-Forced Vibration. Forced-vibration instruments drive specimens at specific frequencies and determine the response, usually over a range of temperatures. Storage and loss moduli or related parameters are determined. Series of modulus-temperature curves can be generated by making measurements at several different frequencies. Because thermal and mechanical transitions are functions of frequency as well as temperature, data from such curves can be used to calculate activation energies of transitions. In addition, frequencies can be chosen to represent or approximate polymer processing effects and use conditions. The considerable interest in simultaneous measurements of mechanical and dielectric properties has led many laboratories to build their own instruments. A successful example of this is the use of vibrating electrodes to monitor complex elastic modulus and dielectric properties (282). Examples of nonresonance-forced vibration instruments include the Imass Dynastat Mechanical Energy Resolver and Rheovibron Autovibron, the TA Instruments 983 DMA and a newer instrument, the DMA 2980, the Metravib Micromechanalyzer, the Polymer Laboratories (now Rheometric Scientific) dynamic mechanical thermal analyzer (DMTA), and the Rheometrics solids analyzer RSA II and RDS II dynamic analyzer.

The TA Instruments 983 DMA is a versatile instrument that can measure a number of viscoelastic properties. It can evaluate materials over a modulus range of 10^{-3} to 2×10^{11} Pa, a temperature range of -150 to 500°C , and at frequencies of 10^{-3} to 10 Hz. Specimen geometries include rectangles, rods, films, and supported materials such as liquid paints and resins on glass cloth. It can also make resonant frequency, creep, and stress-relaxation measurements. TA Instruments introduced in the 1990s a new dynamic mechanical analyzer, the DMA 2980 (Fig. 45). It has a modulus range of 10^3 to 10^{12} Pa, a temperature range of -150 to 600°C , and a frequency range of 10^{-2} to 200 Hz. Many different modes of operation are possible.



Fig. 45. TA Instruments DMA 2980 dynamic mechanical analyzer. Courtesy of TA Instruments, Inc.

The Imass Dynastat (283) is a mechanical spectrometer noted for its rapid response, stable electronics, and exact control over long periods of time. It is capable of making both transient experiments (creep and stress relaxation) and dynamic frequency sweeps with specimen geometries that include tension-compression, three-point flexure, and sandwich shear. The frequency range is 0.01–100 Hz (0.1–200 Hz optional), the temperature range is -150 to 250°C (extendable to 380°C), and the modulus range is 10^4 – 10^{11} Pa.

The Imass Mechanical Energy Resolver (MER) makes clever use of the orientation of an eccentric shaft on a variable speed motor to give a sinusoidal displacement of up to 2-mm amplitude to the specimen. Total force of up to 20 kg is possible, allowing the characterization of materials in the linear and well into the nonlinear viscoelastic region. The mechanical arrangement offers the possibility of unattended long-time operation, ideal for fatigue testing. The MER resolves complex viscoelastic properties by direct measurement of mechanical loss power and complex power. The frequency range is 0.5 to 60 Hz. Temperature can be controlled from -125°C to 205°C .

The Imass Rheovibron is perhaps the earliest and best-known dynamic mechanical spectrometer. It has been widely used in polymer research over the years and, in its automated form, continues to be favored. A simplified schematic of the Rheovibron is shown in Figure 46. The specimen, usually in the form of a rod, fiber, or bar, is attached to strain gauges at each end, one of which (G_1) is a stress transducer, which measures the force applied by the driver; the other (G_2) records the sample deformation. A sinusoidal tensile strain is applied at a given frequency to one end of the sample by the driver. The stress response is measured at the other end. The instrument gives a direct reading of the absolute values of the stress and strain amplitudes and the tangent of the phase angle between stress and strain. From these quantities the dynamic complex E^* , the tensile storage E' , and loss E'' moduli can be calculated. The modulus range is 10^5 – 10^{11} Pa, and the temperature range extends to 200°C ; a special temperature chamber allows a temperature range of 150 to 300°C . The oscillatory frequencies are 3.5, 11, 35, and 110 Hz (0.01–1.0 Hz optional).

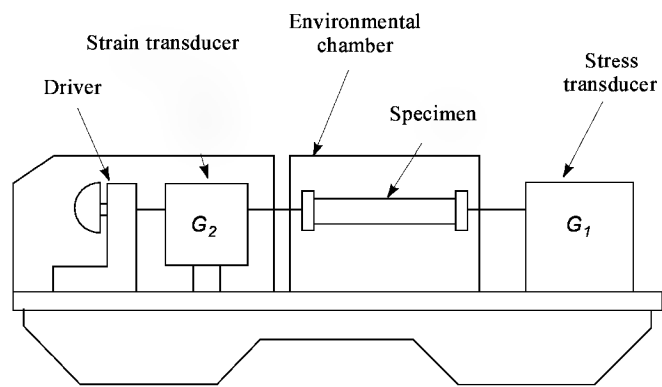


Fig. 46. Schematic diagram of a dynamic mechanical analyzer based on the nonresonance-forced vibration principle (Rheovibron-type).

The Autovibron is the automated version of the Rheovibron (284). It can be purchased in automated form, or the automation system can be added later. Automation improves accuracy and productivity and eliminates the problems and operator dependence of results often associated with the manual instrument. Automation enables the Rheovibron to compete with the other commercial dynamic mechanical instruments.

The Metravib Micromecanalyser is an inverted torsional pendulum, but unlike the torsional pendulums described earlier, it can be operated as a forced-vibration instrument. It is fully computerized and automatically determines G' , G'' , and $\tan \delta$ as a function of temperature at low frequencies (10^{-5} –1 Hz). Stress relaxation and creep measurements are also possible. The temperature range is -170 to 400°C . The Micromecanalyser probably has been used more for the characterization of glasses and metals than for polymers, but has proved useful for determining glassy-state relaxations and microstructures of polymer blends (285) and latex films (286).

The Rheometric Scientific Mk III DMTA (287,288) is a versatile instrument that provides six different measurement modes: cantilever bending, shear sandwich, tension, compression, creep, and a mode of operation mimicking a TMA. Sixteen measurement frequencies are available, ranging from 0.01 to 200 Hz. The temperature range is -150 to 500°C or ambient to 800°C with the high temperature option. Specimens must be in the form of bars or films or be supported by filter paper or a glass braid. Probably the most common modes of operation are the double cantilever and the tensile mode. In double cantilever, the specimen is held at its ends and driven by a vibrating drive shaft which is clamped to the center of the specimen (Fig. 47). In tensile mode the specimen is also clamped at both ends but is driven along the axis between the clamps as one of them is part of the vibrating shaft. As with a number of instruments, several frequencies can be monitored at the same time, as shown in the modulus and $\tan \delta$ plots of Figure 48; the modulus range is 10^3 – 10^{11} N m⁻² (10^4 – 10^{12} dyn cm⁻²). Operation is computer controlled, and cure and embrittlement can be monitored as a function of time as well as of the usual thermal transitions. An optical oven is available that allows uv cure and degradation to be followed.

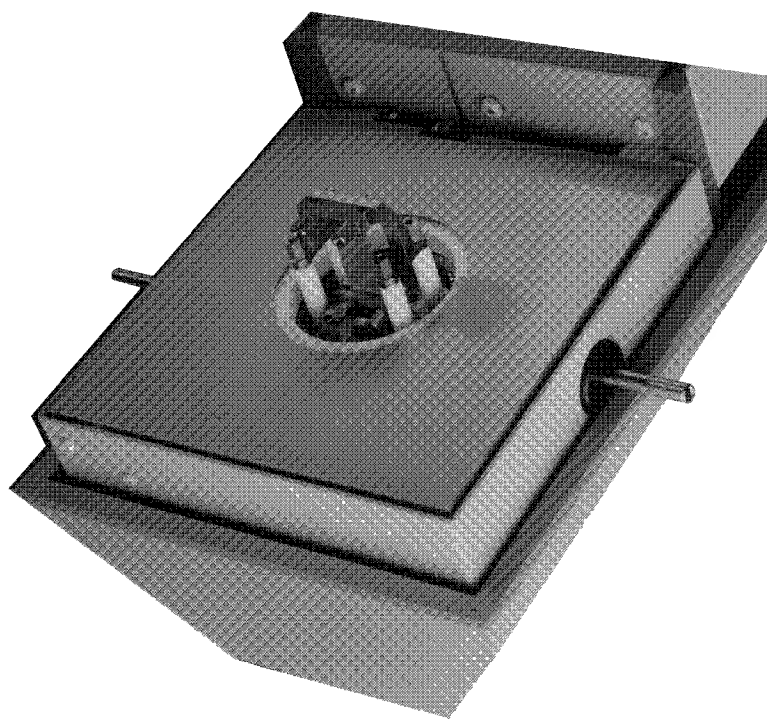


Fig. 47. Tensile specimen holder for DMTA.

Courtesy of Rheometric Scientific, Inc.

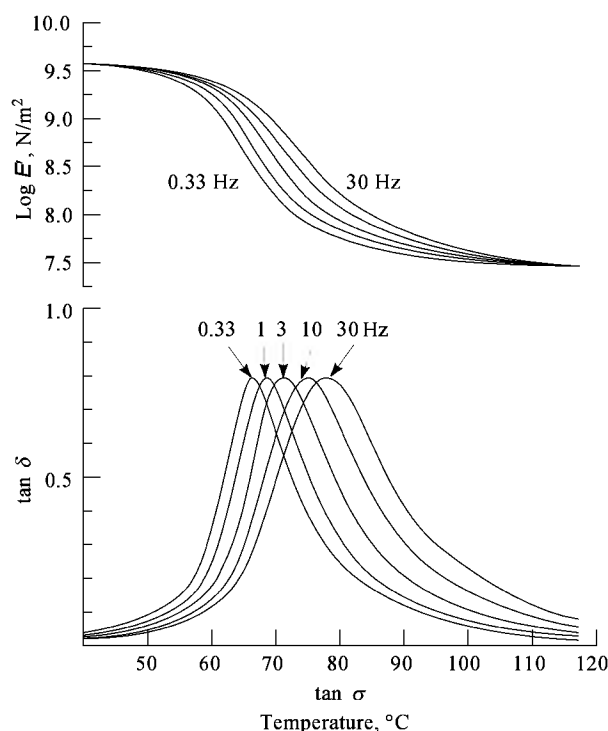


Fig. 48. Multifrequency modulus and $\tan \delta$ vs temperature plots. To convert N m⁻² to psi, multiply by 1.45×10^4 .

The Rheometric Scientific solids analyzer RSA II is a versatile, easy-to-use instrument with a number of test geometries. It can measure dynamic tension of films and fibers, dynamic compression of foams and elastomers, and dual-cantilever bending of polymer films, composites, and ceramics. It can also measure stress relaxation. The instrument has a frequency range of 0.001–100 rad s⁻¹ and a temperature range of -150 to 500°C (extendable to 600°C). Menu-driven software programs the instrument, runs the experiment, and collects the data. The RSA II automatically compensates for thermal expansion and contraction as well as specimen hardening and softening during temperature sweeps. The instrument has a built-in microprocessor that acts as a diagnostics system, which allows rapid identification of instrument problems. A new option allows the rheological and optical properties of transparent materials to be measured simultaneously by making it possible for a light beam to be directed through the specimen.

The Rheometric Scientific RDA II dynamic analyzer is designed for characterization of polymer melts and solids in the form of rectangular bars. It makes computer-controlled measurements of dynamic shear viscosity, elastic modulus, loss modulus, $\tan \delta$, and linear thermal expansion coefficient over a temperature range of ambient to 600°C (-150°C optional) at frequencies 10⁻⁵–500 rad s⁻¹. It is particularly useful for the characterization of materials that experience considerable changes in properties because of thermal transitions or chemical reactions.

Fluids. The previous methods were designed for solid specimens, although some can be used for fluids if a solid support or carrier is used. The fluid must be highly viscoelastic for data to register, and absolute modulus values are difficult to determine because of the presence of the support. Different instrumentation is needed for the determination of meaningful modulus values over wide viscosity and elasticity ranges.

Viscoelastic measurements of dilute solutions are useful because they allow studies of the conformational dynamics of long-chain macromolecules in solution without interference from intermolecular interactions. Such measurements require special instrumentation and techniques. An example is the Birnboim-Schrag multiple-lumped resonator (MLR), which has undergone considerable development and modification since it was first described (289). It measures the changes in frequency and bandwidth of the resonance modes of elements (lumps) on a cylindrical rod undergoing torsional oscillations while surrounded by a solvent or dilute polymer solution. One particular instrument (242,290) has an aluminum rod machined into five lumps. The resonator is clamped at the top and bottom, and oscillation (100–6000 Hz) is induced by driving a magnet that is near the tip. The displacements are small (<0.002") to avoid nonlinear spring effects. Movement is measured with a helium–neon laser optical system; control is by a small computer.

With appropriate calibration the complex characteristic impedance at each resonance frequency can be calculated and related to the complex shear modulus, G^* , of the solution. Extrapolations to zero concentration yield the intrinsic storage and loss moduli [G'] and [G''], respectively, which are molecular properties. In the viscosity range of 0.5–50 mPa·s, the instrument provides valuable experimental data on dilute solutions of random coil (291), branched (292), and rod-like (293) polymers. The upper limit for shearing frequency for the MLR is 800 Hz. High frequency (20 to 500 K Hz) viscoelastic properties can be measured with another instrument, the high frequency torsional rod apparatus (HFTRA) (294).

Viscoelastic fluids that are more concentrated are characterized with devices that are similar to the rotational viscometers described previously. However, instead of constant rotational motion in one direction, a sinusoidal oscillatory motion is provided. Some instruments allow both viscosity and viscoelastic measurements.

Viscoelasticity can also be determined by a controlled stress rheometer. The shape of a creep curve can show that a fluid is viscoelastic, and the amount of recovery after the stress is removed gives a measure of elasticity.

Specific Fluids Viscoelastometers. The ATS RheoSystems Stresstech rheometer can carry out oscillatory measurements over a frequency range of 6.3×10^{-5} to 630 rad s⁻¹. The Bohlin VOR rheometer and the mechanical spectrometer both allow oscillatory measurements. The former has a frequency range of 10⁻⁴–10 Hz, the latter 10⁻⁴–100 Hz. The maximum angular amplitude is 0.02 rad for the VOR and 0.5 rad for the mechanical spectrometer.

The Metravib MVA Mach I (Viscoelasticimeter) and the Mach II (Viscoanalyser) are oscillatory instruments that can be used to characterize the viscoelastic properties of fluids, semisolids, and solids. These instruments are not rotational devices, but are vibrational in nature. Geometries include tensile, shear, annular shear, annular pumping, and three-point bending. The Mach I must be operated manually, but can be linked to a computer. The Mach II is fully automated. The instruments have frequency ranges in the region of 5–10³ Hz. Both allow measurements between -70 and 180°C, with an optional range of -100 to 350°C. The modulus range is 10³–10¹² N m⁻², and the viscosity range is 0.1–10⁶ Pa·s. Metravib is a French acronym for the measurement and treatment of vibrations and noise. Metravib's instruments are widely used for research and problem-solving in these areas.

The Rank Pulse Shearometer is neither a rotational nor an ordinary oscillatory instrument. The shear modulus is determined by measuring the propagation velocity of a small-amplitude torsional shear wave (10,295). The specimen is placed between two stainless steel disks, each of which is coupled to a piezoelectric transducer. The drive disk is oscillated (~250 Hz), and the small angular displacement produces a torsional shear wave, which travels through the specimen and arrives at the receiving disk after some delay. The propagation velocity of the shear wave is determined from the transit time and the distance between the disks. The shear modulus is calculated by $G' = \rho v^2$, where ρ is the density of the specimen and v the propagation velocity. The angular displacement at the drive disk is typically only 10⁻⁴ rad, about two orders of magnitude less than with oscillatory techniques. Therefore, measurements can be made on dispersions and other structured fluids without changing the structure. This is useful for studying thixotropy, sol–gel transitions, and flocculation processes and characterizing paints, inks, and adhesives. The Rank Pulse Shearometer is also used to study polymer lattices

(295–297).

Rheometric Scientific markets several devices designed for characterizing viscoelastic fluids. These instruments measure the response of a liquid to sinusoidal oscillatory motion to determine dynamic viscosity as well as storage and loss moduli. The Rheometric Scientific line includes a fluids spectrometer (RFS-II), a dynamic spectrometer (RDS-7700 series II), and a mechanical spectrometer (RMS-800). The fluids spectrometer is designed for fairly low viscosity materials. The dynamic spectrometer can be used to test solids, melts, and liquids at frequencies from 10^{-3} to 500 rad/s and as a function of strain amplitude and temperature. It is a stripped down version of the extremely versatile mechanical spectrometer, which is both a dynamic viscometer and a dynamic mechanical testing device. The RMS-800 can carry out measurements under rotational shear, oscillatory shear, torsional motion, and tension compression, as well as normal stress measurements. Step strain, creep, and creep recovery modes are also available. It is used on a wide range of materials, including adhesives, pastes, rubber, and plastics.

The RMS-800 provides steady-shear rotational rates from 10^{-6} to 100 rad/s and oscillatory frequencies from 10^{-3} to 100 rad/s. An autotension device compensates for expansion or contraction. With the standard 25- and 50-mm parallel plates, the viscosity range is 50– 10^{11} mPa·s, and the shear modulus range is 8×10^3 to 10^9 N/m². These ranges can be expanded with nonstandard plates, cones, and a Couette system. The temperature range is 20–350°C (–150°C optional).

The TA Instruments CSL² rheometer can perform low frequency oscillatory measurements as well as steady-state viscosity determinations, even though it has a simple mechanical system. The sinusoidal wave form is generated mathematically in the computer rather than with an electromechanical drive system. The stress is controlled, and the resulting strain is determined and stored in memory. The computer analyzes the wave form and calculates the viscosity and elasticity of the specimen at the frequency of the test. As of this writing (1996), the oscillation software covers a frequency range of 10^{-4} –40 Hz. This range could be increased as faster software and computers become available.

The Weissenberg Rheogoniometer is well suited to research on homogeneous viscoelastic fluids and elastic melts. For oscillatory shear a second motor-drive mechanism is added. This allows the use of 60 frequencies in the range of 7.6×10^{-5} to 40 Hz at amplitudes between 2×10^{-3} and 3×10^{-2} rad. An electronic circuit improves the precision of oscillatory measurements, particularly at frequencies near the natural resonance frequency of the instrument itself (298).

Time–Temperature Superposition and Master Curves. Because the modulus of a viscoelastic material varies with time and temperature, measurements must be made over wide ranges of these variables for full characterization. This is often impossible, particularly for very long and very short times. Even if it were possible, the amount of experimental work involved would be prohibitive. Therefore, techniques have been developed to determine modulus values and curves at times or temperatures not attainable experimentally. A series of stress–relaxation, indentation, or modulus curves measured at different temperatures can be shifted on a log time axis to give a single modulus time master curve that covers a wide time range. A creep–compliance–time master curve can be constructed in the same fashion, as can a similar curve by using the families of modulus curves at different frequencies, such as those in Figure 48, generated by dynamic mechanical instruments with multiplexing capabilities. These curves can be shifted to a single modulus–frequency master curve because time and temperature have equivalent effects on modulus and other viscoelastic quantities (145,270,271,276–279).

The shift factor required to superimpose a set of data for an amorphous polymer is described mathematically by the WLF equation (eq. 8) (24), which becomes

$$\log a_T = \frac{17.4(T - T_g)}{51.6 + (T - T_g)}$$

where a_T is the shift factor, the amount of horizontal shift of the time scale. Figure 49 shows the preparation of a modulus master curve from experimentally measured stress–relaxation curves at various temperatures (299).

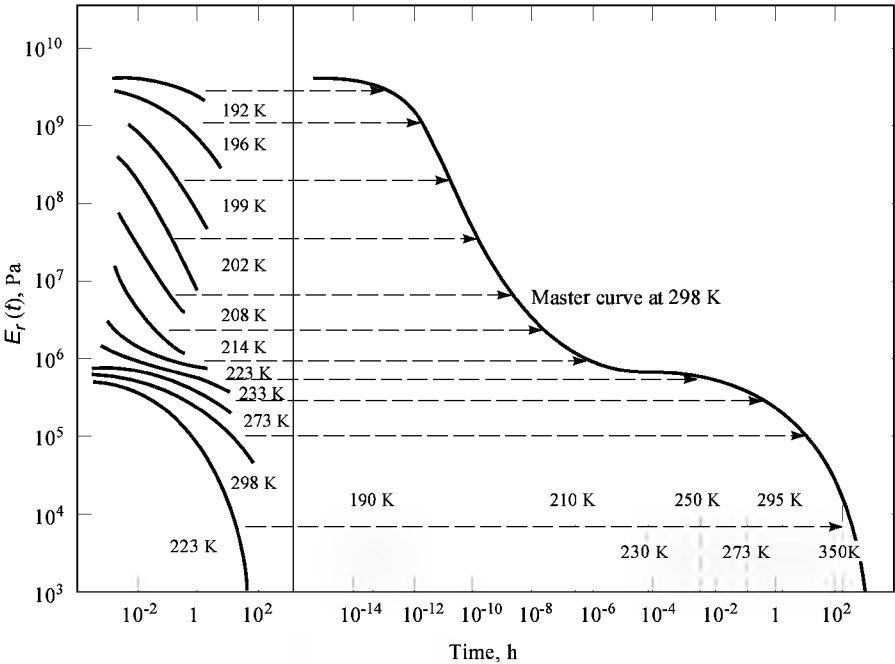


Fig. 49. Illustration of the time–temperature superposition principle as based on stress–relaxation data for polyisobutylene (299,300). To convert Pa to dyn/cm², multiply by 10.

Courtesy of International Textbook Co., Ltd.

Master curves can also be constructed for crystalline polymers, but the shift factor is usually not the same as the one calculated from the WLF equation. An additional vertical shift factor is usually required. This factor is a function of temperature, partly because the modulus changes as the degree of crystallinity changes with temperature. Because crystallinity is dependent on aging and thermal history, vertical factors and crystalline polymer master curves tend to have poor reproducibility.

Master curves can be used to predict creep resistance, embrittlement, and other property changes over time at a given temperature, or the time it takes for the modulus or some other parameter to reach a critical value. For example, a rubber hose may burst or crack if its modulus exceeds a certain level, or an elastomeric mount may fail if creep is excessive. The time it takes to reach the critical value at a given temperature can be deduced from the master curve. Frequency-based master curves can be used to predict impact behavior or the damping ability of materials being considered for sound or vibration deadening. The theory, construction, and use of master curves have been discussed (145,242,271,277,278,299,300).

Practical Rheology

A common problem in rheology is that technologists in charge of research and development, quality control, and problem-solving are rarely trained in rheology, and rheologists are rarely trained in the technology of the product or process. The technologist may use flow to describe effects that are not connected with rheology in the way that a trained rheologist would interpret them. For example, a paint technologist may describe a paint as having poor flow, meaning that the surface has imperfections or defects. However, the same sample may flow easily in a rheological sense during application and cure. The solution of the problem may not involve a change in viscosity, and a small amount of a surfactant may improve the surface appearance of the paint film.

A good working rule in the correlation of viscosity measurements with processing conditions is to select and compare viscosities of materials at the shear rate that corresponds to the process in question. However, problems are encountered with high shear processes. For example, a viscometer with a sufficiently high shear rate may not be available. High shear rates can give heating effects, slip, and erroneous viscosity values. These difficulties can be solved by careful extrapolation from lower shear measurements; Casson plots and other viscosity models can be useful.

Often, a simple ordering of good and poor materials can provide insight into the desired range of viscosity required for satisfactory processing. This strategy assumes that the viscosity of the material is the significant controlling parameter in the process, an assumption that should always be tested. Often, other factors have an equal or even greater influence on the system. For example, in the case of thin films and surfaces, the surface tension and physics of wetting can play a role equal to that of viscosity.

Few processes involve a single shear rate or set of mechanical conditions. Typical processes involve low, intermediate, and high shear regions or sections or stages. For example, a paint or coating may be pumped (intermediate shear rate), be sprayed (high shear rate), coalesce, and flow to form a uniform film (intermediate to low shear rate) and sag or run under gravity (low shear rate) during application. Thus a complete rheological evaluation of a material to determine its processing characteristics requires consideration of the viscosity of the material over an extended shear rate range. A given material may process well at low or intermediate shear rates, but fail completely at higher shear rates.

Additional complications can occur if the mode of deformation of the material in the process differs from that of the measurement method. Most fluid rheology measurements are made under shear. If the material is extended, broken into droplets, or drawn into filaments, the extensional viscosity may be a more appropriate quantity for correlation with performance. This is the case in the parting nip of a roller in which filamenting paint can cause roller spatter if the extensional viscosity exceeds certain limits (109). In a number of cases shear stress is the key factor rather than shear rate, and controlled stress measurements are necessary.

These examples indicate that the correlation of rheological measurements with product and process performance requires study, knowledge, and hard work (3,10,87,146,158,162,163,301). Extensive applications of rheological measurements have been made to food products (20,87,234,236,302–306), lubricants (307–310), adhesives (191–193,311–315), paints (10,18,22,39–41,109,110,150,158,159,164,177,179–184,190,316–318), printing inks (3,159,195,198,319), rubber (301,320–323), polymers and plastics (43–56,145,200–206,240,267–274,276–279,324–327), sealants (328), foams (329–333), slurry fuels (334) and other slurries (335,336), cosmetics (87,337), concentrated suspensions (338–344), crude oil (345,346), powders and granular material (347–350), even rock, ice, snow and other natural geological materials (351–354). Careful rheological measurements and a thorough knowledge of the literature should enable the investigator to meet most rheological challenges.

BIBLIOGRAPHY

"Rheology" in *ECT* 1st ed., Vol. 11, pp. 730–748, by J. R. Van Wazer, Monsanto Chemical Co.; "Viscometry" in *ECT* 1st ed., Vol. 14, pp. 756–776, by F. H. Stross and P. E. Porter, Shell Development Corp.; "Rheology" in *ECT* 2nd ed., Vol. 17, pp. 423–444, by J. W. Lyons, Monsanto Co.; "Viscometry" in *ECT* 2nd ed., Vol. 21, pp. 460–484, by R. S. Stearns, Sun Oil Co.; "Rheological Measurements" in *ECT* 3rd ed., Vol. 20, pp. 259–319, by P. E. Pierce and C. K. Schöff, PPG Industries, Inc.

1. M. M. Couette, *Ann. Chim. Phys. Ser. VI*, **21**, 433 (1890).
2. R. J. Donnelly, *Phys. Today* **44**(11), 32 (Nov. 1991).
3. H. Green, *Industrial Rheology and Rheological Structures*, John Wiley & Sons, Inc., New York, 1949.
4. M. Van Dyke, *An Album of Fluid Motion*, The Parabolic Press, Stanford, Calif., 1982.
5. D. V. Boger and K. Walters, *Rheological Phenomena in Focus*, Elsevier Science Publishing Co., Inc., New York, 1993.
6. H. A. Barnes and K. Walters, *Rheol. Acta*, **24**, 323 (1985).
7. C. W. Macosko, in R. J. Goldstein, ed., *Fluid Mechanics Measurements*, Hemisphere Publishing Corp., New York, 1983, p. 423.
8. T. C. Papanastasiou, *J. Rheol.* **31**, 385 (1987).
9. E. A. Collins, C. H. Chen, and J. P. Padolewski, *J. Rheol.* **32**, 163 (1988).
10. R. Eley, in J. Koleske, ed., *Paint Testing Manual: Gardner-Sward Handbook*, 14th ed., ASTM, Philadelphia, 1995, p. 333.
11. J. M. Dealy, *J. Rheol.* **39**, 253 (1995).
12. M. Rosen, *Polym. Plast. Technol. Eng.* **12**(1), 1 (1979).
13. N. Casson, in C. C. Mills, ed., *Rheology of Disperse Systems*, Pergamon Press, New York, 1959, p. 84.
14. W. K. Asbeck, *Off. Dig. Fed. Soc. Paint Technol.* **33**(432), 65 (1961).
15. R. V. Williamson, *Ind. Eng. Chem.* **21**(11), (1929).
16. M. M. Cross, *J. Colloid Sci.* **20**, 417 (1965).
17. W. H. Bauer and E. A. Collins, in F. R. Eirich, ed., *Rheology*, Vol. 4, Academic Press, Inc., New York, 1967, p. 423.
18. O. C. C. Lin, *J. Appl. Polym. Sci.* **19**, 199 (1975).
19. R. Roscoe, in J. J. Hermans, ed., *Flow Properties of Disperse Systems*, North-Holland Publishing Co., Amsterdam, the Netherlands, 1953, p. 1.
20. G. W. Scott Blair, *Elementary Rheology*, Academic Press, Inc., New York, 1969.
21. J. R. Van Wazer and co-workers, *Viscosity and Flow Measurement, A Laboratory Handbook of Rheology*, Wiley-Interscience, New York, 1963.
22. P. E. Pierce, *J. Paint Technol.* **43**(557), 35 (1971).
23. *Ibid.*, **41**(533), 383 (1969).
24. M. L. Williams, R. F. Landel, and J. D. Ferry, *J. Am. Chem. Soc.* **77**, 3701 (1955).
25. Z. W. Wicks, Jr. and co-workers, *J. Coat. Technol.* **57**(725), 51 (1985).
26. Z. W. Wicks, Jr., F. N. Jones, and S. P. Pappas, *Organic Coatings: Science and Technology*, Vol. II, John Wiley & Sons, New York, 1994, p. 13.
27. O. F. Solomon and I. Ciuta, *J. Appl. Polym. Sci.* **6**, 683 (1962).
28. K. K. Chee, *J. Appl. Polym. Sci.* **34**, 891 (1987).
29. M. Kurata and co-workers, in J. Brandrup and E. H. Immergut, eds., *Polymer Handbook*, 2nd ed., Wiley-Interscience, New York, 1975, p. IV-1.
30. S. H. Aharoni, *J. Appl. Polym. Sci.* **21**, 1323 (1977).
31. J. F. Rabek, *Experimental Methods in Polymer Chemistry*, Wiley-Interscience, New York, 1980.
32. M. Bohdanecky and J. Kovar, *Viscosity of Polymer Solutions*, Elsevier, Amsterdam, the Netherlands, 1982.
33. V. N. Kalashnikov, *J. Rheol.* **38**, 1385 (1994).
34. G. V. Vinogradov and A. Ya. Malkin, *Rheology of Polymers*, Springer-Verlag, New York, 1980.
35. T. G. Fox, S. Gratch, and S. Loshaek, in F. R. Eirich, ed., *Rheology*, Vol. I, Academic Press, Inc., New York, 1956, p. 431.
36. G. Berry and T. G. Fox, *Adv. Polym. Sci.* **5**, 261 (1968).
37. H. Fujita and Y. Einaga, *Rheol. Acta*, **25**, 487 (1986).

38. W. W. Graessley, in J. E. Mark, ed., *Physical Properties of Polymers*, American Chemical Society, Washington, D.C., 1984, Chapt. 3, p. 97.
39. C. K. Schoff, *Prog. Org. Coat.* **4**, 189 (1976); *Am. Chem. Soc. Div. Polym. Mater. Sci. Eng. Prepr.* **55**, 8 (1986); *Proceedings of the 14th Waterborne and Higher-Solids Coatings Symposium*, New Orleans, La., Feb. 1987, p. 252.
40. M. A. Sherwin, J. V. Koleski, and R. A. Taller, *J. Coatings Technol.* **53**(683), 35 (1981).
41. R. Buter, in G. D. Parfitt and A. V. Patsis, eds., *Proceedings of the Fifth International Conference on Organic Coatings Science and Technology*, Athens, Greece, 1979, Technomics Publishing Co., Westport, Conn., 1981, p. 12.
42. D. R. Paul, J. E. St. Lawrence, and J. H. Troell, *Polym. Eng. Sci.* **10**, 70 (1970).
43. F. N. Cogswell, *Polymer Melt Rheology*, Halsted Press, a division of John Wiley & Sons, Inc., New York, 1981, p. 40.
44. G. Pezzin, *Materie Plastiche ed Elastomeri*, Aug. 1962; translated and reprinted as *Application Series, PC-12*, Instron Corp., Canton, Mass., 1972.
45. R. A. Mendelson, in N. M. Bikales, ed., *Encyclopedia of Polymer Science and Technology*, Vol. 8, Wiley-Interscience, New York, 1968, pp. 587–620.
46. J. L. White, in K. Walters, ed., *Rheometry: Industrial Applications*, Research Studies Press, a division of John Wiley & Sons, Inc., New York, 1980, pp. 209–280.
47. J. M. Dealy and K. F. Wissbrun, *Melt Rheology and Its Role in Plastics Processing*, Van Nostrand Reinhold Co., Inc., New York, 1990.
48. J. Meissner, *Makromol. Chem. Macromol. Symp.* **56**, 25 (1992).
49. J. M. Dealy, *Rheometers for Molten Plastics*, Society of Plastics Engineers and Van Nostrand Reinhold Co., Inc., New York, 1982; J. M. Dealy and T. O. Broadhead, *Polym. Eng. Sci.* **33**, 1513 (1993).
50. W. Gleissle, *Rheology*, **4**, 13 (1994).
51. R. Hingmann and B. L. Marzinke, *J. Rheol.* **38**, 573 (1994).
52. B. Maxwell and M. Nguyen, *Polym. Eng. Sci.* **19**, 1140 (1979).
53. B. Maxwell, in C. D. Craver, ed., *Polymer Characterization*, American Chemical Society, Washington, D.C., 1983, p. 149.
54. P. J. Hansen and M. C. Williams, *Polym. Eng. Sci.* **27**, 586 (1987).
55. H. T. Pham and E. Meinecke, *J. Appl. Polym. Sci.* **53**, 257, 265 (1994).
56. J. G. Oakley and A. J. Giacomini, *Polym. Eng. Sci.* **34**, 580 (1994).
57. S. F. Edwards, *Proc. Phys. Soc. London*, **92**, 9 (1967); P. Doi and S. Edwards, *J. Chem. Soc. Faraday Trans. 2*, **74**(10), 1789 (1978).
58. P.-G. de Gennes, *J. Chem. Phys.* **55**, 572 (1971); *Physics Today* **36**(6), 33 (June 1983).
59. J. Maddox, *Nature*, **330**, 11 (1987).
60. J. des Cloizeaux, *Macromolec.* **23**, 4678 (1990).
61. W. W. Graessley, *Accounts Chem. Res.* **10**, 332 (1977).
62. S. H. Wasserman and W. W. Graessley, *J. Rheol.* **36**, 543 (1992).
63. W. H. Tuminello, *Polym. Eng. Sci.* **26**, 1339 (1986); A. Ya. Malkin and A. E. Teishev, *Polym. Eng. Sci.* **31**, 1590 (1992).
64. S. H. Wasserman, *J. Rheol.* **39**, 601 (1995).
65. M. I. Aranguren and co-workers, *J. Rheol.* **36**, 1165 (1992).
66. R. A. Mendelson, *Polym. Eng. Sci.* **8**(3), 235 (1968); **16**(10), 690 (1976).
67. P. D. Driscoll and D. G. Bogue, *J. Appl. Polym. Sci.* **39**, 1755 (1990).
68. S. E. Kadijk and B. H. A. A. Van Den Brule, *Polym. Eng. Sci.* **34**, 1535 (1994).
69. H. L. Goldsmith and S. G. Mason, in F. R. Eirich, ed., *Rheology*, Vol. 4, Academic Press, Inc., New York, 1967, p. 232.
70. D. Lerche, B. Koch, and G. Vlastos, *Rheology*, **3**, 105 (1993).
71. H. L. Frisch and R. Simha, in Ref. 35, p. 525.
72. I. M. Krieger, in R. Buscall, T. Corner, and I. F. Stageman, eds., *Polymer Colloids*, Elsevier Science Publishing Co., Inc., New York, 1986.
73. J. Mewis and P. D'Haene, *Macromol. Chem. Macromol. Symp.* **68**, 213 (1993).
74. J. Mewis, in G. Astarita and co-workers, eds., *Rheology*, Vol. 1, Plenum Press, New York, 1980, p. 149.
75. A. B. Metzner, *J. Rheol.* **29**, 739 (1985).
76. W. B. Russel, *J. Rheol.* **24**, 287 (1980).
77. S. C. Tsai, D. Botts, and J. Plouff, *J. Rheol.* **36**, 1291 (1992).
78. Lj. Djakovir, I. Sefer, and T. Djakovir, *J. Disp. Sci. Technol.* **13**, 697 (1992).
79. W. B. Russel and P. R. Sperry, *Progr. Org. Coatings*, **23**, 305 (1994).
80. T. Shikata and D. S. Pearson, *J. Rheol.* **38**, 601 (1994).
81. J. Persello and co-workers, *J. Rheol.* **38**, 1845 (1994).
82. N. Phan-Thien, *J. Rheol.* **39**, 679 (1995).
83. T. A. Strivens, *Colloid Polym. Sci.* **261**, 74 (1983).
84. E. Windhab, *Rheology*, **2**, 102 (1992).
85. T. Tadros, *Chem. Ind. (London)*, 907 (Dec. 7, 1992).
86. A. L. Zydney, J. D. Oliver III, and C. K. Colton, *J. Rheol.* **35**, 1639 (1991).
87. P. Sherman, *Industrial Rheology*, Academic Press, Inc., New York, 1970.
88. D. Quemeda, *Rheol. Acta*, **16**, 82 (1977).
89. N. L. Ackermann and H. T. Shen, *AIChE J.* **25**, 237 (1979).
90. M. Mooney, *J. Colloid Sci.* **6**, 162 (1951).
91. D. C.-H. Cheng, *Chem. Ind. (London)*, 408 (May 17, 1980).
92. I. M. Krieger and T. J. Dougherty, *Trans. Soc. Rheol.* **3**, 137 (1959).
93. T. A. Witten and M. F. Cates, *Science*, **232**, 1607 (1986).
94. L. M. Sander, *Nature*, **322**, 789 (1986).
95. T. A. Strivens, *J. Colloid Interface Sci.* **57**, 476 (1976).
96. W. H. Boersma, J. Laven, and H. N. Stein, *AIChE J.* **36**, 321 (1990); *J. Rheol.* **39**, 841 (1995).
97. E. P. Vrahopoulou and A. J. McHugh, *J. Non-Newt. Fluid Mech.* **25**, 157 (1987).
98. G. I. Taylor, *Proc. R. Soc. London Ser. A*, **138**, 41 (1932).
99. Ref. 87, p. 137.
100. J. W. Hill and J. A. Cuculo, *J. Macromol. Sci. Rev. Macromol. Chem.* **C14**, 107 (1976).
101. J. Ferguson and N. E. Hudson, in R. A. Pethrick, ed., *Polymer Yearbook 2*, Harwood Academic Publishers, London, 1985, p. 155.
102. J. M. Dealy, *Polym. Eng. Sci.* **11**, 433 (1971).
103. J. Meissner, *Polym. Eng. Sci.* **27**, 537 (1987).
104. J. Ferguson and K. Walters, *Chem. Britain*, **24**(1), 39 (1988).
105. K. Walters, in P. Moldenaers and R. Keunings, eds., *Theor. Appl. Rheol.*, Vol. 1, Elsevier, Amsterdam, the Netherlands, 1992, p. 16.
106. J. Ferguson and N. E. Hudson, *Int. J. Polym. Mat.* **20**(3–4), 181 (1993).
107. M. Takahashi and co-workers, *J. Rheol.* **37**, 827 (1993).
108. C. D. Denson, *Polym. Eng. Sci.* **13**, 125 (1973).
109. J. E. Glass, *J. Coatings Technol.* **50**(641), 56, 72 (1978); D. A. Soules, R. H. Fernando, and J. E. Glass, *J. Rheol.* **32**, 181, 199 (1988).
110. D. F. Massouda, *J. Coatings Technol.* **57**(722), 27 (1985).
111. H. M. Laun and H. Munstedt, *Rheol. Acta*, **17**, 415 (1978).

112. K. F. Wissbrun, *Chemtracts: Macromol. Chem.* **1**(2), 131 (1990).
113. S. Catrei, C. W. Macosko, and H. H. Winter, *J. Rheol.* **25**, 433 (1981).
114. J. L. White, *Appl. Polym. Symp.* **33**, 31 (1978).
115. W. Winslow, *J. Appl. Phys.* **20**, 1137 (1949).
116. T. C. Halsey, *Science*, **258**, 761 (1992).
117. L. Marshall, C. F. Zukoski, and J. Goodwin, *J. Chem. Soc. Faraday Trans. I*, **85**, 2785 (1989).
118. R. T. Bonnecaze and J.-F. Brady, *J. Rheol.* **36**, 73 (1992).
119. T. C. Halsey, J. E. Martin, and D. Adolf, *Phys. Rev. Lett.* **68**, 1519 (1992).
120. D. R. Gamota, A. S. Wineman, and F. E. Filisko, *J. Rheol.* **37**, 919 (1993).
121. H. Aldersey-Williams, *New Scientist*, **125**(178), 37 (Mar. 17, 1990).
122. N. Webb, *Chem. Britain*, **26**(4), 338 (1990).
123. T. C. B. McLeish, T. Jordan, and M. T. Shaw, *J. Rheol.* **35**, 427 (1991).
124. D. L. Hontsock, R. F. Novak, and G. J. Chaundy, *J. Rheol.* **35**, 1305 (1991).
125. W. S. Yen and P. J. Achorn, *J. Rheol.* **35**, 1375 (1991).
126. H. Conrad and co-workers, *J. Rheol.* **35**, 1393 (1991).
127. W. A. Bullough, *Endeavor*, **14**(4), 165 (1991).
128. T. C. Jordan, M. T. Shaw, and T. C. B. McLeish, *J. Rheol.* **36**, 441 (1992).
129. F. E. Filisko, *Chem. Ind. (London)*, (10), 370 (May 18, 1992); *J. Rheol.* **34**, 539 (1990); *J. Rheol.* **35**, 399 (1991).
130. H. Block, *Chemtech*, **22**(6), 368 (1992).
131. Z. Lon and R. D. Ervin, *J. Rheol.* **37**, 55 (1993).
132. K. M. Blackwood and H. Block, *Trends Polym. Sci.* **1**(4), 98 (1993).
133. C. F. Zukoski, *Annu. Rev. Mater. Sci.* **23**, 45 (1993).
134. T. C. Halsey, *Adv. Mater.* **5**(10), 711 (1993).
135. E. Westkamper and J. Meschke, *Rheology*, **3**, 243 (1993).
136. Y. Otsubo and K. Edamura, *J. Rheol.* **38**, 1721 (1994).
137. B. Abu-Jdayil and P. O. Brunn, *Rheology*, **4**, 187 (1994).
138. A. Inoue and S. Maniwa, *J. Appl. Polym. Sci.* **55**, 113 (1995).
139. H. Conrad, Y. Li, and Y. Chen, *J. Rheol.* **39**, 1041 (1995).
140. *J. Rheol.* **35**(7), 1303–1461 (1991).
141. J. Rabinow, *AIEE Trans.* **67**, 1308 (1948).
142. H. Janocha and B. Rech, *Rheology*, **3**, 39 (1993); **4**, 198 (1994).
143. C. Zener, *Elasticity and Anelasticity of Metals*, University of Chicago Press, Chicago, Ill., 1948.
144. S. P. Timoshenko and J. N. Goodier, *Theory of Elasticity*, 3rd ed., McGraw-Hill Book Co., Inc., New York, 1969.
145. J. R. Fried, *Polymer Science and Technology*, Prentice Hall Inc., Englewood Cliffs, N.J., 1995.
146. J. Ferguson and Z. Kemblowski, *Applied Fluid Rheology*, Elsevier Science Publishing Co., Inc., New York, 1991.
147. Y. P. Khanna and K. R. Slusarz, *Polym. Eng. Sci.* **33**, 122 (1993).
148. J. K. Gillham and C. K. Schoff, *J. Appl. Polym. Sci.* **20**, 1875 (1976).
149. J. Starita, in G. Astarita and co-workers, eds., *Rheology*, Vol. 2, Plenum Press, New York, 1980, p. 229; *Rheometrics System Four*, Rheometrics, Inc., Union, N.J., 1980; J. Meissner, *J. Appl. Polym. Sci.* **16**, 2877 (1972), Weissenberg Rheogoniometer data.
150. O. C. C. Lin, *Chemtech*, **5**(1), 51 (1975).
151. A. S. Lodge, T. S. R. Al-Hadithi, and K. Walters, *Rheol. Acta*, **26**, 516 (1987).
152. A. Ait-Kadi, L. Choplin, and P. J. Carreau, *Polymer Eng. Sci.* **29**, 1265 (1989).
153. G. A. Nucez and co-workers, *J. Rheol.* **38**, 1251 (1994).
154. D. Hadjistamov, *Rheology*, **5**, 29 (1995).
155. E. B. Bagley, *J. Appl. Phys.* **28**, 624 (1957).
156. R. J. Gardner and P. C. Senanayake, *Rev. Sci. Instrum.* **57**, 3129 (1986).
157. M. A. Haney, *J. Appl. Polym. Sci.* **30**, 3023 (1985).
158. T. C. Patton, *Paint Flow and Pigment Dispersion*, 2nd ed., John Wiley & Sons, Inc., New York, 1979.
159. M Euverard, *The Efflux Type Viscosity Cup*, Scientific Section, National Paint, Varnish & Lacquer Association, Washington, D.C., 1948; *ASTM Bull.* **162**, 67 (Oct. 1950).
160. W. Gleissle, *Rheology*, **4**, 138 (1994).
161. D. Hadjistamov and F. Gutekunst, *Rheology*, **4**, 29 (1994).
162. K. Walters, ed., *Rheometry: Industrial Applications*, Research Studies Press, a division of John Wiley & Sons, Inc., New York, 1980.
163. R. W. Whorlow, *Rheological Techniques*, Halsted Press, a division of John Wiley & Sons, Inc., New York, 1980.
164. T. A. Strivens, in R. Lambourne, ed., *Paint and Surface Coatings*, Ellis Horwood Ltd., Chichester, U.K., 1987, chapt. 14, p. 547.
165. P. Toepke, *Rheology*, **1**, 102 (1991).
166. M. Breuker, *Rheology*, **1**, 179 (1991).
167. R. E. Smith, *J. Rheol.* **28**, 155 (1984).
168. S. S. Davis, J. J. Deer, and B. J. Warburton, *Phys. Educ.* **2**, 1 (1968); *J. Pharm. Pharmacol.* **20**, 836 (1968).
169. T. E. R. Jones, J. M. Davies, and H. A. Barnes, in B. Mena and co-workers, eds., *Proceedings of the IX International Congress of Rheology*, Vol. 4, Univ. Nacional Autonoma de Mexico, Mexico City, Mexico, 1984, p. 45.
170. L. Bohlin, *News Bohlin Rheol.* (2), 6 (1986).
171. A. J. P. Franck, *J. Rheol.* **29**, 833 (1985); *Proceedings of the XI International Congress on Rheology*, Brussels, Belgium, Aug. 1992, p. 982.
172. M. Schmidt and A. J. P. Franck, *Rheology*, **4**, 23 (1994).
173. T. S. Wilson, *Appl. Rheol. Meas. Newsletter*, **3**(1), 3 (Mar. 1994).
174. Q. D. Nguyen and D. V. Boger, *J. Rheol.* **27**, 331 (1983); *J. Rheol.* **29**, 335 (1985); *Rheol. Acta*, **26**, 508 (1987).
175. A. S. Yoshimure and co-workers, *J. Rheol.* **31**, 699 (1987).
176. A. M. Kraynik and co-workers, in Ref. 169, p. 77.
177. T. C. Patton, *J. Paint Technol.* **38**(502), 656 (1966); *Cereal Sci. Today*, **14**, 178 (1969).
178. P. A. Sanford and co-workers, *J. Appl. Polym. Sci.* **22**, 701 (1978).
179. R. E. Smith, *J. Coatings Technol.* **54**(694), 21 (1982).
180. J. H. Monk, *J. Oil Colour Chem. Assoc.* **49**, 543 (1966).
181. P. S. Pond and C. J. H. Monk, *J. Oil Colour Chem. Assoc.* **53**, 876 (1970).
182. M. L. Colclough, N. D. P. Smith, and T. A. Wright, *J. Oil Colour Chem. Assoc.* **63**, 183 (1980).
183. D. M. Gans, *J. Paint Technol.* **44**(571), 68 (1972).
184. A. Quach and P. E. Pierce, *J. Paint Technol.* **45**(586), 69 (1973).
185. L. A. Mondy and co-workers, *Int. J. Multiphase Flow*, **12**, 497 (1986).
186. A. L. Graham and co-workers, *Appl. Phys. Lett.* **50**, 127 (1987).

187. B. J. Briscoe, P. F. Luckham, and S. R. Ren, *Colloids Surfaces*, **62**, 153 (1992).

188. A. Quach and C. M. Hansen, *J. Paint Technol.* **46**(592), 40 (1974).

189. W. Goring, N. Dingerdissen, and C. Hartmann, *Farbe Lack*, **83**, 270 (1977).

190. C. K. Schoff, *Rheology*, Federation of Societies for Coatings Technology, Blue Bell, Pa., 1991, p. 38; *Proceedings of the 5th Annual ESD Adv. Coatings Conference*, Dearborn, Mich., Nov. 1995, p. 77.

191. J. Johnston, *Adhes. Age*, **26**, 34 (1983).

192. F. Urushizuki, H. Yamaguchi, and H. Mizumachi, *J. Adhes. Soc. Jpn.* **20**, 195 (1984).

193. H. Mizumachi and T. Saito, *J. Adhes.* **20**, 83 (1986).

194. N. A. Park and T. F. Irvine, Jr., *Waerme Stoffuebertrag* **18**, 201 (1984); *Rev. Sci. Instrum.* **59**, 2051 (1988); *Amer. Lab.* **21**(12), 8 (Dec. 1989).

195. R. Bassemir, *Am. Inkemaker*, **39**(4), 24 (1961).

196. J. Mewis, in Ref. 162, p. 324.

197. G. Pangalos and J. M. Dealy, *J. Oil Colour Chem. Assoc.* **68**(3), 59 (1985).

198. G. Pangalos, J. M. Dealy, and M. B. Lyre, *J. Rheol.* **29**, 471 (1985).

199. K. K. Chee, K. Sato, and A. Rudin, *J. Appl. Polym. Sci.* **20**, 1467 (1976).

200. J. M. Dealy and S. S. Soong, *J. Rheol.* **28**, 355 (1984).

201. S. R. Doshi and J. M. Dealy, *J. Rheol.* **31**, 563 (1987).

202. J. M. Dealy and A. J. Giacomin, in A. A. Collyer and D. W. Clegg, eds., *Rheological Measurement*, Elsevier Applied Science, New York, 1988.

203. A. J. Giacomin, T. Samurkas, and J. M. Dealy, *Polym. Eng. Sci.* **29**, 499 (1989).

204. N. R. Demarquette and J. M. Dealy, *J. Rheol.* **36**, 1007 (1992).

205. S. G. Hatzikiriakos and J. M. Dealy, *J. Rheol.* **35**, 497 (1991); *Int. Polym. Process*, **8**, 36 (1993).

206. I. B. Kazatchov, S. G. Hatzikiriakos, and C. W. Stewart, *Polym. Eng. Sci.* **35**, 1864 (1995).

207. J. D. Ferry, *Meas. Control*, **11**, 89 (Sep. Oct. 1977).

208. H. M. Laun and H. Schuch, *J. Rheol.* **33**, 119 (1989).

209. J. Ferguson and M. K. H. El-Tawashi, in Ref. 149, p. 257.

210. C. A. Moore and J. R. A. Pearson, *Rheol. Acta*, **14**, 436 (1975).

211. K. M. Baid and A. B. Metzner, *Trans. Soc. Rheol.* **21**, 237 (1977).

212. D. M. Jones, K. Walters, and P. R. Williams, *Rheol. Acta*, **26**, 20 (1987).

213. F. D. Martischius, *Rheol. Acta*, **21**, 288, 311 (1982).

214. K. K. K. Chao and M. C. Williams, *J. Rheol.* **27**, 451 (1983).

215. V. Tirtaatmadja and T. Sridhar, *J. Rheol.* **37**, 1081 (1993).

216. G. G. Fuller and co-workers, *J. Rheol.* **31**, 235 (1987).

217. M. R. Anklam, G. G. Warr, and R. K. Prud'homme, *J. Rheol.* **38**, 797 (1994); R. K. Prud'homme and G. G. Warr, *Langmuir*, **10**, 3419 (1994).

218. R. L. Ballman, *Rheol. Acta*, **4**, 137 (1965).

219. F. N. Cogswell, *Rheol. Acta*, **8**, 187 (1969).

220. J. Meissner, *Rheol. Acta*, **8**, 78 (1969).

221. H. Munstedt, *J. Rheol.* **20**, 211 (1976).

222. *Ibid.*, **23**, 421 (1979).

223. A. Franck and J. Meissner, *Rheol. Acta*, **23**, 117 (1984).

224. C. D. Denson and J. R. Gallo, *Polym. Eng. Sci.* **11**, 174 (1971).

225. D. D. Joye, G. W. Poehlein, and C. D. Denson, *Trans. Soc. Rheol.* **16**, 421 (1972).

226. J. M. Maerker and W. R. Schowalter, *Rheol. Acta*, **13**, 627 (1974).

227. J. Meissner, T. Raible, and S. E. Stephenson, *J. Rheol.* **25**, 1, 673 (1981).

228. S. H. Chatraei, C. W. Macosko, and H. H. Winter, *J. Rheol.* **25**, 433 (1981).

229. A. C. Papanastasiou, L. E. Scriven, and C. W. Macosko, *J. Rheol.* **27**, 387 (1983).

230. P. R. Soskey and H. H. Winter, *J. Rheol.* **29**, 493 (1985).

231. S. A. Kahn, R. K. Prud-homme, and R. G. Larson, *Rheol. Acta*, **26**, 144 (1987).

232. T. C. Hsu and I. R. Harrison, *Polym. Eng. Sci.* **31**, 223 (1991).

233. M. Takahashi and co-workers, *J. Rheol.* **37**, 827 (1993).

234. H. Huang and J. L. Kokini, *J. Rheol.* **37**, 879 (1993).

235. H. C. Kim, A. Pendse, and J. R. Collier, *J. Rheol.* **38**, 831 (1994).

236. K. Wikström, L. Bohlin, and A.-C. Eliasson, *Rheology*, **4**, 192 (1994).

237. N. Golshan Ebrahimi and co-workers, *Acta Polym.* **46**, 267 (1995); *J. Rheol.* **39**, 1385 (1995).

238. R. K. Gupta and T. Sridhar, in A. A. Collyer and D. W. Clegg, eds., *Rheological Measurement*, Elsevier Applied Science, New York, 1988, p. 211; T. Sridhar, in Y. L. Yeow and P. H. T. Uhlherr, eds., *Proceedings of the 5th National Conference on Rheology*, Australia Society of Rheology, Melbourne, Australia, 1990, p. 1.

239. B. H. Bersted, *Polym. Eng. Sci.* **33**, 1079 (1993).

240. J. Meissner, *Rheology*, **5**, 120 (1995).

241. P. E. Pierce, in R. R. Myers and J. S. Long, eds., *Treatise on Coatings*, Vol. 2, Marcel Dekker, Inc., New York, 1969, pt. I, p. 99.

242. J. D. Ferry, *Viscoelastic Properties of Polymers*, 2nd ed., John Wiley & Sons, Inc., New York, 1970.

243. J. C. Monpagens and co-workers, *J. Polym. Sci. Pol. Phys. Ed.* **15**, 767 (1977).

244. A. Matthiesen and co-workers, *Amer. Lab.* **23**(1), 44 (Jan. 1991).

245. C. M. Neag and co-workers, *J. Coatings Technol.* **65**(826), 37 (1993).

246. J. P. Ibar and co-workers, in *Polymer Characterization: Physical Property, Spectroscopic, and Chromatographic Methods*, American Chemical Society, Washington, D.C., 1990, p. 167.

247. Technical data, Thermold Partners L.P., Glenbrook Industrial Park, Stamford, Conn.

248. H. Bangert, A. Wagendristel, and H. Aschinger, *Colloid Polymer Sci.* **259**, 238 (1981).

249. A. Zosel, *Farbe Lack*, **82**, 15 (1976).

250. *Ibid.*, **80**, 1130 (1974); **83**, 804 (1977).

251. A. Zosel, *Progr. Org. Coat.* **8**, 47 (1980).

252. A. G. Epprecht, *Farbe Lack*, **80**, 505 (1974); **82**, 685 (1976).

253. R. L. J. Morris, *J. Oil Colour Chem. Assoc.* **53**, 761 (1970).

254. K. T. Gillen, *J. Appl. Polym. Sci.* **22**, 1291 (1978).

255. D. J. Skrovanek and C. K. Schoff, *Progr. Org. Coat.* **16**, 135 (1988).

256. C. K. Schoff and P. Kamarchik, in A. T. Riga and C. M. Neag, eds., *Materials Characterization by Thermomechanical Analysis*, ASTM STP 1136, ASTM Philadelphia, Pa., 1991, p. 138.

257. G. M. Pharr and W. C. Oliver, *Mat. Res. Soc. Bulletin*, **17**, 28 (1992).

258. W. C. Oliver, *Mat. Res. Soc. Bulletin* **11**, 15 (1986).

259. M. F. Doerner and W. D. Nix, *J. Mater. Res.* **1**, 601 (1986).

260. W. C. Oliver and G. M. Pharr, *J. Mater. Res.* **7**, 1564 (1992).
261. K. T. Gillen, R. L. Clough, and C. A. Quintana, *Polym. Degrad. Stab.* **17**, 31 (1987).
262. L. W. Hill, *Progr. Org. Coatings*, **5**, 277 (1977).
263. A. Toussaint and L. D'Hont, *J. Oil Col. Chem. Assoc.* **64**, 302 (1981).
264. L. W. Hill, *Mechanical Properties of Coatings*, Federation of Societies for Coatings Technology, Philadelphia, Pa., 1987.
265. Z. W. Wicks, Jr., F. N. Jones, and S. P. Pappas, *Organic Coatings Science and Technology*, Vol. II, John Wiley & Sons, New York, 1994, chapt. 24, p. 105.
266. L. W. Hill, in J. V. Koleske, ed., *Paint Testing Manual: Gardner-Sward Handbook*, 14th ed., ASTM, Philadelphia, 1995, p. 534.
267. R. E. Challis and co-workers, *J. Appl. Polym. Sci.* **44**, 65 (1992).
268. A. W. Chow and J. L. Bellin, *Polym. Eng. Sci.* **32**, 182 (1992).
269. J. B. Enns and J. K. Gillham, *Computer Applications in Applied Polymer Science*, American Chemical Society, Washington, D.C., 1982, chapt. 20.
270. T. Alfrey, Jr., *Mechanical Behavior of High Polymers*, Wiley-Interscience, New York, 1984.
271. J. J. Aklonis, W. J. MacKnight, and M. Shen, *Introduction to Polymer Viscoelasticity*, Wiley-Interscience, New York, 1972.
272. R. F. Boyer, *Am. Chem. Soc. Org. Coat. Plast. Div. Prepr.* **44**, 491 (1981).
273. J. K. Gillham, *AIChE J.* **20**, 1066 (1974).
274. J. K. Gillham, *Developments in Polymer Characterization-3*, Applied Science Publishers, London, 1982, chapt. 5.
275. S. Ikeda, *Progr. Org. Coat.* **1**, 205 (1972–1973).
276. N. G. McCrum, B. E. Read, and G. Williams, *Anelastic and Dielectric Effects in Polymer Solids*, John Wiley & Sons, Inc., New York, 1967.
277. L. E. Nielson, *Mechanical Properties of Polymers and Composites*, Vols. 1 and 2, Marcel Dekker, Inc., New York, 1974.
278. A. V. Tobolsky, *Properties and Structures of Polymers*, John Wiley & Sons, Inc., New York, 1960.
279. I. M. Ward, *Mechanical Properties of Solid Polymers*, John Wiley & Sons, Inc., New York, 1971.
280. C. K. Schoff, *J. Coatings Technol.* **49**(633), 62 (1977).
281. J. K. Gillham, *Polym. Eng. Sci.* **26**, 1429 (1986).
282. B. Binder and co-workers, *Colloid Polymer Sci.* **271**, 22 (1993).
283. S. S. Sternstein, in C. D. Craver, ed., *Polymer Characterization*, American Chemical Society, Washington, D.C., 1983, p. 123.
284. S. M. Webber, J. A. Manson, and R. W. Lang, in Ref. 283, p. 109.
285. J. Y. Cavaille and co-workers, *Polymer*, **27**, 549, 606 (1986).
286. *Ibid.*, p. 693.
287. R. E. Wetton, in Ref. 283, p. 27.
288. R. E. Wetton and co-workers, *Amer. Lab.* **25**(1), 15 (Jan. 1993).
289. J. L. Schrag and R. M. Johnson, *Rev. Sci. Instrum.* **42**(2), 224 (1971).
290. E. J. Amis and co-workers, *Anal. Chim. Acta*, **189**, 199 (1986); *Macromolecules*, **22**, 4528 (1989); *Rev. Sci. Instrum.* **60**, 2780 (1989).
291. K. Osaki, J. L. Schrag, and J. D. Ferry, *Macromolecules*, **5**, 144 (1972).
292. Y. Mitsuda and co-workers, *Polym. J.* **4**, 24 (1972).
293. N. Nemoto and co-workers, *Biopolymers*, **14**, 409 (1975).
294. T. M. Stokich and co-workers, *J. Rheol.* **38**, 1195 (1994).
295. R. Buscall and co-workers, *J. Chem. Soc. Faraday Trans. 1*, **78**, 2873 (1982).
296. *Ibid.*, p. 2889.
297. J. Goodwin and co-workers, *J. Colloid Interface Sci.* **97**(2), 488 (1984).
298. L. Conde, M. A. Rubio, and E. Riande, *J. Rheol.* **31**, 527 (1987).
299. J. M. G. Cowie, *Polymers: Chemistry and Physics of Modern Materials*, International Textbook Co., Ltd., Aylesbury, U.K., 1973, p. 236.
300. E. Catsiff and A. V. Tobolsky, *J. Colloid Sci.* **10**, 375 (1955).
301. J. W. White, *Rubber Chem. Technol.* **42**, 257 (1960).
302. C. Gallegos and M. Berjano, *J. Rheol.* **36**, 465 (1992).
303. K. Seethamraju and M. Bhattacharya, *J. Rheol.* **38**, 1029 (1994).
304. D. Jaros and H. Rohm, *Rheology*, **4**, 77 (1994).
305. H.-D. Tscheuschner, *Rheology*, **4**, 83 (1994).
306. V. Breton and co-workers, *Rheology*, **5**, 24 (1995).
307. E. Kuhn, *Rheology*, **2**, 252 (1992).
308. P. M. Cann and H. A. Spikes, *Lubric. Eng.* **48**, 335 (1992).
309. R. Mas and A. Magnin, *J. Rheol.* **38**, 889 (1994).
310. E. Kuhn, *Rheology*, **4**, 129 (1994).
311. P. R. Steiner, *J. Adhesion*, **16**, 279 (1984).
312. G. Schmitt, J. Wiley, and J. Gotro, *Polym. Eng. Sci.* **29**, 329 (1989).
313. S. Vleeshouwers, A. M. Jamieson, and R. Simha, *Polym. Eng. Sci.* **29**, 662 (1989).
314. M. H. Pahl and D. Heskamp, *Rheology*, **3**, 97 (1993).
315. N. R. Manoj and co-workers, *J. Adhesion Sci. Technol.* **7**, 363 (1993).
316. M. Breucker, *Rheology*, **3**, 48 (1993).
317. M. Osterhold and Y. Jäger, *Rheology*, **3**, 250 (1993).
318. N. Leskovsek, L. Tusar, and M. Tusar, *Rheology*, **5**, 140 (1995).
319. Y.-H. Zang and co-workers, *J. Rheol.* **35**, 345 (1991); T. Hartford, *Am. Inkemaker*, **72**(11), 42 (1994).
320. Y.-H. Zang, R. Muller, and D. Froelich, *J. Rheol.* **30**, 1165 (1986).
321. J. L. White and co-workers, *J. Rheol.* **35**, 167 (1991).
322. M. H. Wagner and J. Schaeffer, *J. Rheol.* **37**, 643 (1993).
323. A. Roychoudhury and S. K. De, *J. Appl. Polymer Sci.* **50**, 811 (1993).
324. D. J. Plazek, *Macromolecules*, **16**, 1469 (1983); *Polym. Eng. Sci.* **24**, 111 (1984); *J. Rheol.* **36**, 1671 (1992).
325. T. S. Chow, *J. Mater. Sci.* **25**, 957 (1990); *Polymer*, **32**, 29 (1991); *J. Rheol.* **36**, 1707 (1992); *Adv. Polym. Sci.* **103**, 149 (1992).
326. P. J. Halley and M. E. Mackay, *J. Rheol.* **38**, 1 (1994); *High Perform. Polym.* **6**, 405 (1994).
327. G. Sanz, *J. Appl. Polym. Sci.* **55**, 75 (1995).
328. E. A. Collins, C. H. Chen, and J. P. Padolewski, *J. Rheol.* **32**, 163 (1988); **36**, 117, 131 (1992).
329. A. M. Kraynik and M. G. Hansen, *J. Rheol.* **31**, 175 (1987).
330. D. Weaire and T.-L. Fu, *J. Rheol.* **32**, 271 (1988).
331. E. Mora, L. D. Artavia, and C. W. Macosko, *J. Rheol.* **35**, 921 (1991).
332. A. M. Kraynik, D. A. Reinelt, and H. M. Prince, *J. Rheol.* **35**, 1235 (1991).
333. T. Okuzono, K. Kawasaki, and T. Nagai, *J. Rheol.* **37**, 571 (1993).
334. S.-F. Lin and R. S. Brodkey, *J. Rheol.* **29**, 147 (1985).
335. R. J. Mannheimer, *J. Rheol.* **35**, 113 (1991).
336. M. K. Agarwala, B. R. Patterson, and P. E. Clark, *J. Rheol.* **36**, 319 (1992).

337. D. Holland, *Rheology*, **2**, 108 (1991).
338. A. J. Poslinski and co-workers, *J. Rheol.* **32**, 703 (1988).
339. Y.-S. Lin, T.-J. Liu, and N.-J. Chu, *J. Appl. Polym. Sci.* **42**, 1767 (1991).
340. S. Priel and G. Torriano, *Rheology*, **1**, 223 (1991).
341. S. C. Tsai, D. Botts, and J. Plouff, *J. Rheol.* **36**, 1291 (1992).
342. J. Schröder, *Rheology*, **2**, 40 (1992).
343. S. A. Khan and N. J. Zoeller, *J. Rheol.* **37**, 1225 (1993).
344. D. M. Kalyon, *Chemtech*, **25**(5), 22 (May 1995).
345. L. T. Wardhaugh and D. V. Boger, *AIChE J.* **37**, 871 (1991).
346. L. E. Sanchez and J. L. Zakin, *Ind. Eng. Chem. Res.* **33**, 3256 (1994).
347. R. Jackson, *J. Rheol.* **30**, 907 (1986).
348. P. K. Haff, *J. Rheol.* **30**, 931 (1986).
349. R. Ocone and G. Astarita, *J. Rheol.* **37**, 727 (1993).
350. Y. Tomita and co-workers, *J. Rheol.* **38**, 231 (1994).
351. W. R. Peltier, in A. Dziewonski and E. Boschi, eds., *Physics of the Earth's Interior*, North Holland, Amsterdam, the Netherlands, 1980, p. 362.
352. Symposium on Rheology of Geological Materials, *J. Rheol.* **28**(6), 663–798 (1984).
353. P. England, *Nature* **381**, 23 (1996).
354. C. H. Jones, J. R. Unruh, and L. J. Sonder, *Nature* **381**, 37 (1996).

General References

Refs. 10, 12, 20, 21, 34, 43, 49, 87, 145, 146, 150, 158, 162–164, 242, 270, 271, 276–279, and 299 are good general references.
Ref. 21, although out of date, is the single most useful reference in this area. Ref. 49 is a good reference for viscometers in general.
H. A. Barnes, J. F. Hutton, and K. Walters, *An Introduction to Rheology*, Elsevier Applied Science, New York, 1989.
J. M. Burgers, J. J. Hermans, and G. W. Scott Blair, eds., *Deformation and Flow*, Vols. I–VII, Interscience Publishers, New York, 1952–1956.
B. D. Coleman, H. Markovitz, and W. Noll, *Viscometric Flow of Non-Newtonian Fluids*, Springer-Verlag, New York, 1966.
A. A. Collyer, ed., *Techniques in Rheological Measurement*, Chapman and Hall, New York, 1993.
A. A. Collyer and D. W. Clegg, eds., *Rheological Measurement*, Elsevier Applied Science, New York, 1988.
A. Dinsdale and R. Moore, *Viscometry and Its Measurement*, The Institute of Physics and the Physical Society, Reinhold Publishing Corp., New York, 1963.
F. R. Eirich, ed., *Rheology: Theory and Application*, Vols. 1–5, Academic Press, Inc., New York, 1956–1970.
A. G. Fredrickson, *Principles and Applications of Rheology*, Prentice-Hall, Inc., Englewood Cliffs, N.J., 1964.
C. W. Macosko, ed., *Rheology: Principles, Measurements, and Applications*, VCH Publishers, Inc., New York, 1994.
M. Reiner, *Deformation, Strain and Flow*, 2nd ed., Interscience Publishers, New York, 1960.
M. Rosen, *Polym. Plast. Technol. Eng.* **12**(1), (1979). A very useful, practical review.
S. Ross and I. D. Morrison, *Colloidal Systems and Interfaces*, John Wiley & Sons, New York, 1988.
W. R. Schowalter, *Mechanics of Non-Newtonian Fluids*, Pergamon Press, New York, 1977.
K. Walters, *Rheometry*, Halsted Press, a division of John Wiley & Sons, Inc., New York, 1975.
There are a number of excellent journals that cover or include rheological studies. Two particularly useful ones are the *Journal of Rheology* (free with membership in the Society of Rheology) and the German English *Rheology* (Vincentz Verlag) which provides a lively combination of papers, new equipment descriptions, abstracts of patents and papers, and other information. Other useful journals include *Rheologica Acta*, *Polymer Engineering and Science*, and the *Journal of Applied Polymer Science*.

Clifford K. Schoff
Peter Kamarchik, Jr.
PPG Industries, Inc.

RIBOFLAVIN.

See VITAMINS.

RISK ASSESSMENT.

See HAZARD ANALYSIS AND RISK ASSESSMENT.

RODENTICIDES.

See PESTICIDES.

ROOFING MATERIALS

Roofs are a basic element of shelter from inclement weather. Natural or hewn caves, including those of snow or ice, are early evidence of human endeavors for protection from the cold, wind, rain, and sun. Nomadic people, before the benefits of agriculture had been discovered and housing schemes developed, depended on the availability of natural materials to construct shelters. Portable shelters, eg, tents, probably appeared early in history. Later, more permanent structures were developed from stone and brick. Salient features depended strongly on the availability of natural materials. The Babylonians used mud to form bricks and tiles that could be bonded with mortars or natural bitumen. Ancient buildings in Egypt were characterized by massive walls of stone and closely spaced columns that carried stone lintels to support a flat roof, often made of stone slabs.

As larger and larger structures were desired, building design became an important element in construction. Roofs evolved from a simple covering to roofing systems designed to perform a number of functions to separate indoor and outdoor environments. Selection of roof design depended not only on factors of economy and comfort, but also on the availability of materials and structural and aesthetic factors. Thus, a modern design normally includes a structure to carry loads, insulation to control heat flow, a barrier to control air and vapor flow, and a roofing element to prevent water penetration.

For low slope commercial roofing, bituminous-based roof coverings are the most common systems in the United States. Asphalt-based materials predominate over coal-tar based materials in these systems. For residential roofing materials, various types of roofing products, including asphalt, wood, and tile, are used for both new construction and reroofing.

Continuous or Jointed Systems

Built-Up Roofing. Built-up roofing (BUR) is a continuous-membrane covering manufactured on-site from alternate layers of bitumen, bitumen-saturated or coated felts, or asphalt-impregnated glass mats and surfacings. These membranes are generally applied with hot bitumens or cold applied bituminous adhesives (qv).

The deck may be nailable, eg, wood or light weight concrete, or not, eg, steel or structural concrete. The felts or mats may be organic (cellulose), or fiber glass. The roof slope ranges from dead level (0–2.1 cm m), to flat (2.1–12.5 cm m), to steep (12.5–25 cm m).

Specifications and Application Rates. Specifications as published in catalogs of roofing manufacturers or contractors' associations appear complicated because of the many variations possible. Application methods depend on the type and slope of the deck, the types of insulation and roofing membrane, and the fastening method (Table 1). Built-up roofs generally are not applied to slopes steeper than 25 cm m because application of hot asphalt (qv) to such roofs is difficult, whereas other forms of roofing, such as shingles, are easily applied.

Table 1. Simplified Hot Applied Membrane Material Specifications

Type	ASTM Specification	Reference
organic ply felt	D226, Types I and II	1
base sheet	D2626	2
fiber glass mats	D2178, Types III–VI	3
base sheet	D4601, Types I and II	4
venting base sheet	D4897, Types I and II	5
surfacings		
organic cap sheet	D371	
fiber glass sheet	D3909	6
asphalt	D312, Types I–IV	7
gravel or slag	D1863	8
asphalt-based aluminum coating ^a	D2824, Types I or III	9
emulsified asphalts ^a	D1227, Types II–IV	10
fibered roof coating ^a	D4479	11

^a Asbestos-free.

Common BUR systems are installed in different ways: membrane adhered attached to deck without insulation; insulation adhered attached to deck with membrane applied to insulation; base sheet adhered attached to deck, insulation to base sheet, and top membrane to insulation; and membrane adhered attached to deck and insulation applied over the membrane, the so-called protected-membrane roof.

The components must be anchored as protection against wind uplift, slippage, and membrane movement. Application rates for BUR membranes are given in Table 2. Membrane strength is related to felt or glass-mat strengths and the number of plies. Continuous bonding is inadvisable if the elastic

limit of the membrane will be exceeded, which results in the development of splits in the membrane and leads to roof leaks. Nailing or spot adhesion of the first layer provides relief from concentrated strains.

Table 2. Typical Application Rates for BUR Membranes

Material	ASTM specification	Cold applied, L m ^{2a}	Hot applied, kg m ^{2b}	Reference
primer	D41	0.3		12
bitumen ply	D312, D450		1.22	7, 13
adhesive	D312, D450, D3019	0.4–0.8		7, 13, 14
bitumen surfacing	D312, D450, D1227	0.8–1.6	2.9	7, 10, 13
gravel covering	D1863		19.5	8
slag covering	D1863		14.6	8
asphalt–liquid surfacing cements		0.6–1.22		7, 9–11

^a To convert L m² to gal 100 ft², multiply by 2.45.

^b To convert kg m² to lb 100 ft², multiply by 20.5.

Properties of Ply Felts and Asphalt-Saturated and Coated Felts. The properties of asphalt-saturated and coated felts and impregnated fiber glass mats are given in Table 3. In addition, ASTM D4897 (5) describes an asphalt impregnated and coated glass fiber base sheet that has mineral surfacing on the top and coarse mineral granules on the bottom for use as the first ply of a roofing membrane. The base sheet provides for the lateral release of pressure that might be caused by entrapped vapor.

Table 3. ASTM Requirements for Asphalt-Saturated and Coated Felts for BUR

Property	Organic base sheet ^{a,b}	ASTM D4601 ^c		ASTM D4897 ^d	
		Type I	Type II	Type I	Type II
mass, min, kg m ^{2e}	1.8	0.65	0.75	2.44	2.68
mass desaturated mat, min, g m ^{2e}	253	68	83	73	83
mass % mineral matter, max, %	60	70	70	60	60
passing a 212-μm (No. 70) sieve					
breaking strength, kN m ^f					
with fiber grain	6.1	3.9	7.7	3.9	7.7
across fiber grain	3.5	3.9	7.7	3.9	7.7
permeance, max, ng (Pa·s·m ²) ^g	17				
pliability radius, 13 mm, 10 pass, at 23.1°C		pass	pass		

^c Perforations provide vents for gases liberated during applications^b (4,10).

^d Asphalt-coated glass fiber venting base sheet with fine mineral surfacing on the top side and coarse granules on the bottom side; perforated embossed or not. The coarse granules provide an open, porous channel in the horizontal plane^b (5,11).

^a First ply or as vapor barrier under insulation. Coated on both sides with asphalt and surfaced on top side with fine mineral granules (2,9).

^b Product does not crack, tear, or cause damage upon being unrolled at 10–60°C.

^e To convert g m² to lb 100 ft², multiply by 0.0205.

^f To convert kN m to lbf in., divide by 0.175.

^g To convert ng (Pa·s·m²) to grain (h·in·Hg·ft²), divide by 5.72.

Table 4 describes properties of ASTM ply felts used for BUR. For additional details, the appropriate ASTM specifications should be consulted (1,3). Periodic revisions are made by ASTM committees to keep up with current practices.

Table 4. Properties of Roofing and Waterproofing Ply Felts^a for BUR

Property	Organic felt, asphalt-saturated ^b		Fiber glass mat, asphalt-impregnated ^c		
	Type I	Type II	Type III	Type IV	Type VI
nominal weight, g m ^{2d}	560	1270	474	303	303
felt or mat mass, min, g m ^{2d}	254	488	73	83	93
saturant asphalt mass, min, g m ^{2d}	303	732	308	146	146
breaking strength, min, kN m ^e					
with grain	5.25	7.0	3.85	7.70	10.5
across grain	2.63	3.5	3.85	7.70	10.5
pliability radius, mm, pass, at 25°C	12.7	19.1	13	13	13
ash, %	10	10	70–88	70–88	70–88
loss on heating at 105°C, 5 h max, %	4	4			

^a All are nonsticking upon being unrolled at 10–60°C.

^b Perforated or nonperforated (1).

^c Ref. 3.

^d To convert g m² to lb 100 ft², multiply by 0.0205.

^e To convert kN m to lbf in., divide by 0.175.

Cold Applied Coatings/Adhesives. Cold applied BUR applications do not require heating to fluidize the bitumen on the job. Simple application and economical maintenance are primary considerations. Bitumens are liquefied by dissolving in a solvent (cutbacks) or dispersing in water

(emulsion). The base cutback can be modified by addition of other components, commonly mineral fillers and fibers, to form specialized coatings (qv). At ordinary temperatures, the solvent or water evaporates and forms an adhesive bond.

The compositions of cold applied coatings adhesives are illustrated in Figure 1. These coatings range from thin fluids used for priming deck surfaces to viscous mineral-filled compositions used as flashing cements. The ASTM specifications in References 7 and 9–r11 should be consulted for details (see COATINGS; COATING PROCESSES). Lap cements used with asphalt-roll roofing are specified by ASTM D3019 Type III (14). This specification covers cements consisting of asphalt and petroleum solvent without asbestos. It is an asbestos-free cement.

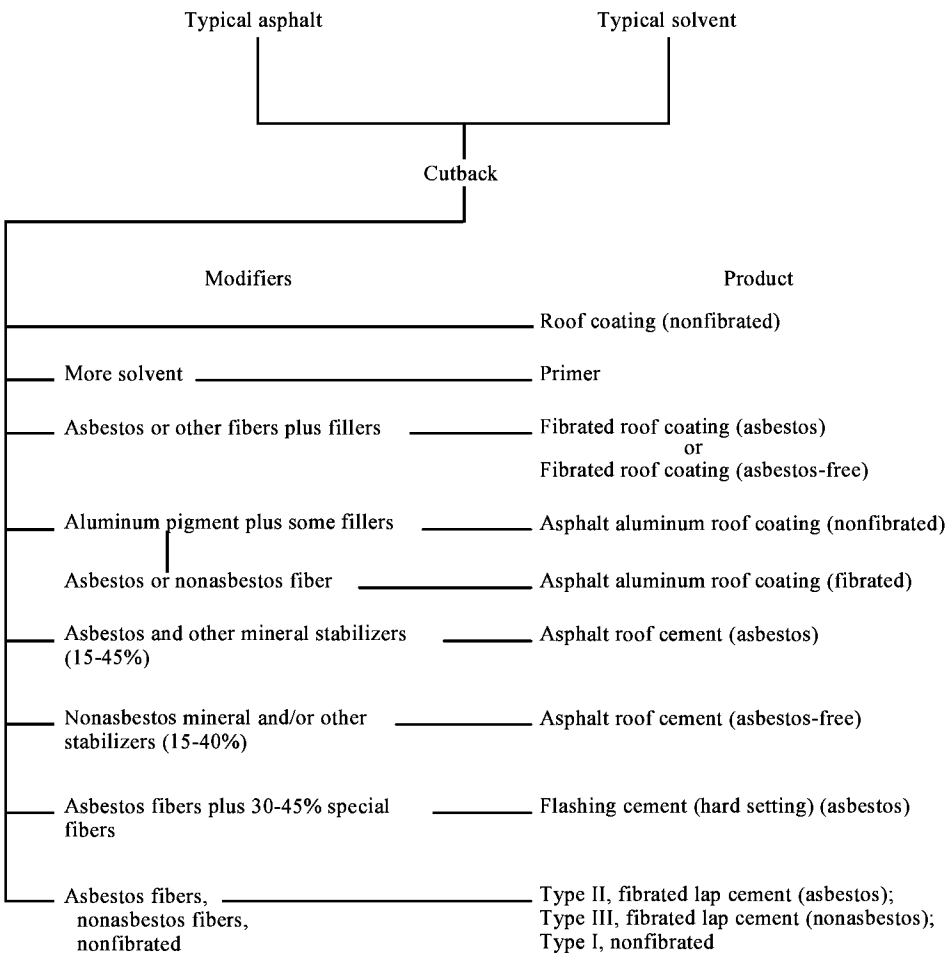


Fig. 1. Production of asphaltic coatings, primers, and cements. Typical asphalt has 18–100 mm penetration at 25°C; typical solvent has a boiling range of 148–210°C. The cutback represents 55–75% asphalt.

Emulsified asphalt used as a protective coating is specified by ASTM D1227 (10). These emulsions are applied above freezing by brush, mop, or spray, and bond to either damp or dry surfaces. Such application is not recommended for inclines <4° to avoid the accumulation of water. However, curing by water evaporation can be slow, and these emulsions may remain water-susceptible.

Protected-Membrane Roofs. Primitive roofs covered with earth and sod over sloping wood decks shingled with bark were early examples of protected-membrane roofs (PMRs). Grass and earth provided insulation and protected the shingled deck from inclement weather.

In modern PMR construction, thermal insulation that is unaffected by water or that can be kept dry in some manner is required. Extruded polystyrene (XEPS) foam insulation boards are commonly employed (see INSULATION, THERMAL). They are placed on top of the waterproofing roof membrane, which is next to the deck. The insulation should not be adhered to the membrane. Ballast at the rate of >48.8 kg m² (1000 lb 100 ft²) holds the insulation in place and offers protection from the sun. The insulation joints are open and drainage must be provided. Various other materials, eg, patio blocks and concrete slabs, are also used as surfacings and ballast. The extra weight imposes more exacting requirements on construction.

Asphalt Roofing Components. Asphalt (qv) is a unique building material which occurs both naturally and as a by-product of crude-oil refining. Because the chemical composition of crude oils differs from source to source, the physical properties of asphalts derived from various crudes also differ. However, these properties can be tailored by further processing to fit the application for which the asphalt will be used. Softening point, ductility, flash point, and viscosity–temperature relationship are only a few of the asphalt properties that are important in the fabrication of roofing products.

Asphalt intended for roofing can be tailored to perform two separate functions. The first is to saturate the organic fiber-based material. This requires that the asphalt be very fluid at processing temperatures so that it can impregnate the base material completely. The second is to coat the saturated roofing to serve as the medium for adhering mineral granule surfacing to the roofing. In the manufacture of roofing on a fiber glass base material, the saturation step is eliminated.

When produced at the refinery, asphalt is soft and sticky and is referred to as flux. Saturants and coating asphalts are both made from the same flux by an oxidizing or dehydrogenation process known as air blowing. During this process, air is bubbled through hot flux. Heat and oxygen cause chemical reactions which increase the molecular weight and carbon:hydrogen ratio, resulting in changed physical properties. Catalysts may be used to produce saturants or coatings that have higher penetration for a given softening point. The blowing process is continually monitored and is completed when the desired properties are produced. The processed asphalt is then pumped to a storage tank prior to delivery to the roofing production line.

At the roofing plant, coating asphalts are blended with mineral stabilizer such as finely ground limestone, slate, flyash, or traprock. The stabilizer increases the coating asphalt's resistance to fire and foot-traffic and adds durability.

Coal-Tar Pitch. Coal tar, a by-product from the coking process of coal for steel production, is also used as a hot-melt adhesive in BURs (see TAR AND PITCH). When used as ply adhesives, coal tars are generally limited to slopes of 2 cm m for organic membranes and <1 cm m for glass fiber systems. The permissible slopes for asphalts without nailing range up to 50 cm m, depending on the asphalt grade specified. Properties of hot applied bitumens for use in BUR as specified by ASTM are given in Tables 5 and 6.

Table 5. Properties of Roofing Asphalt^a

Property	Type I	Type II	Type III	Type IV
----------	--------	---------	----------	---------

flash point ^b , COC ^c , °C, min	246	246	246	246
softening point, °C (range)	57–66	70–80	85–96	99–107
slope guideline, max, cm m	<2.08	<8.33	<25	<50

^a Ref. 7.
^b As of late 1996, ASTM subcommittee D08.03 is considering changing the flash point from 246°C to 260°C.
^c COC = Cleveland open cup.

Table 6. Coal-Tar Pitch Used in Roofing, Dampproofing, and Waterproofing^a

Property	Type I	Type II	Type III
flash point, COC ^b , °C, min	190	175	205
softening point, ring and ball, °C, max	80	80	

^a Ref. 13.
^b COC = Cleveland open cup.

Organic Felts. Roofing felts consist of fiber mats impregnated and or coated with bitumen. Earlier felts contained rag fiber. Extensive weathering tests and field experience demonstrated that Asplund wood-fiber recycled paper felts are satisfactory for BUR and prepared roofing without the addition of variable rag fibers. Organic felts are formed by preparing a water slurry of pulp, after which the fibers are picked up by a rotating screened drum or by a traveling screen. The formed sheet is dried by vacuum suction and steam-heated drums and then calendered and wound into large rolls for subsequent trimming and slitting to the proper width. The organic felt is then saturated by dipping the felt in hot bitumen and passing it over a series of rolls for cooling. The felt is then perforated, marked with guidelines for application, and wound into rolls of convenient size. The small perforations let air and moisture escape during application. Furthermore, the saturant reduces the rate of water vapor absorption. Vapor permeability and water absorption are further reduced by asphalt coatings, stabilized with mineral fillers (qv). A parting agent, eg, sand, talc, or a soap film, is spread on the coating to prevent sticking (see RELEASE AGENTS).

Fiber Glass Felts. Fiber glass felts are manufactured from glass filaments prepared from a molten mixture of silica sodium carbonate, recycled glass, sand, and other ingredients. The filaments may be short (chopped) or continuous. They may be deposited dry or from a dilute water slurry on a screen conveyor. In either process, a binder is applied. Heat is then applied to set the binder. The formed fiber glass mat is then coated with hot asphalt, surfaced with a parting agent (liquid or mineral), cooled, marked with guidelines, and wound into rolls of the desired length. Excellent dimensional stability and low moisture absorption are derived from fiber glass roofing felts. For the production of mineral-surfaced roofing products or shingles, mineral granules are applied to a heavier-coated top side of the felt. Variations in the manufacturing process also produce venting base sheets and flashings. Fiber glass mats are also used to produce modified bitumen roofing products.

Single-Ply Roofing Membranes. The single-ply roofing membrane category derives its name from the installation technique where a single-ply sheet membrane is applied over the roofing deck substrate, giving continuous water-tight coverage (15–31). Single-ply has grown from its beginning in the early 1960s to approximately 45% of the low sloped roofing market. There are approximately 16 manufacturers and an additional 20 marketers of these membranes. These flexible membranes offer light weight, excellent chemical and weather resistance, ease of application and repair, and multiple colors. They can be installed in a number of ways, including the traditional fully adhered, the mechanically fastened, and the ballasted roofing systems. They have been designed to comply with various wind and fire code requirements of the construction industry. Because of the light weight of the adhered and mechanically fastened systems, they sometimes can be installed directly over an existing roof system, thus eliminating the waste that would normally be generated from a roof tear-off. Proper precautions should be taken to remove saturated areas of the old roof before overlaying the new system.

As the single-ply industry grew, its influence in the marketplace has increased. The trade association SPRI (Sheet Membrane and Component Suppliers to the Commercial Roofing Industry) was founded to represent the industry on various technical items and to help educate the industry on single-ply. The single-ply industry has also actively pursued the development of standards for the various single-ply products through ASTM and ANSI. These standards cover areas from the performance of a membrane under aging conditions and the endurance of metal components in corrosive environments, to system designs for wind resistance. The standards for membrane materials provide information comparing unaged material with materials after aging for conditions such as ultraviolet radiation, heat aging, water absorption, ozone (qv), cold temperature flexibility, and linear dimensional changes.

There are two basic categories that these flexible membranes fit into: thermoset and thermoplastic. The first is an elastomeric material that undergoes or has undergone a chemical reaction by the action of heat, catalysts, ultraviolet light, etc, leading to a state where the material is infusible and its shape cannot be altered. This chemical reaction, called curing, can occur in the plant or on the roof. The second is a flexible polymeric material that can be softened repeatedly by heating and hardened by cooling through a temperature range characteristic of the plastic. This material in the softened state can be formed or heat-fused to itself.

Single-ply membranes offer the widest range of systems in the roofing industry. The three basic systems are ballasted, fully adhered, and mechanically fastened. From a cost standpoint, the fully adhered system is the most expensive to install, the ballasted system the least. The protected-membrane roofing system can be used with any of the basic systems. The specifications for these systems are published by the various manufacturers. The following gives a brief description of the roof assemblies.

Ballasted. A ballasted roof assembly consists of a membrane or membrane and substrate material (insulation, slip sheet, etc) loosely laid over a deck with the assembly held in place using ballast. A minimum ballast weight of 48.9 kg m² or 10 pounds per square foot (PSF) is used. The ballast can consist of smooth rounded stone, crushed stone (a separator sheet must be used between the crushed stone and the membrane), or pavers (both standard and lightweight). Both stone and pavers come in a wide variety of colors. The membrane is affixed to the building only at the deck perimeter (roof edge) and at various penetrations. Wall and penetration flashings are typically fully adhered and sealed to prevent water entry into the roof assembly. The maximum slope a ballasted system should be installed over is 16.7 cm m.

Because the ballasted roof uses only 48.8–97.6 kg m² loading, it cannot be evaluated by standard wind uplift test devices such as Factory Mutual's 4470 uplift table or Underwriters' Laboratory's UL 1897 uplift test. Because both devices use air pressure in excess of 146.4 kg m² (30 PSF), the loose ballast would roll off as the membrane inflates. To aid the industry in designing ballasted roof systems for wind, a wind design standard was developed: the ANSI Standard SPRI RP-4: Wind Design Standard for Ballasted Single-Ply Roofing Systems. The standard was developed through wind tunnel work and actual field data documenting the performance of these systems through various wind events, including hurricanes.

Fully Adhered. The substrate, ie, insulation, cover board, etc, that the single-ply membrane is to be attached to is either fully adhered or mechanically fastened to the deck. However, there are also applications where the membrane is adhered directly to the deck. The membrane is then adhered to the substrate. The typical method for adhering the membrane to the substrate is by applying a contact adhesive to the membrane and substrate, rolling the membrane into place, and brooming once the adhesive is ready. There are one-sided applications where the membrane is rolled directly into the adhesive that has been applied to the substrate only. The membrane used in this application method may be fleece-backed. Fully adhered systems can be installed on any slope. The fully adhered application offers a smooth surface that is easy to maintain and inspect, as well as excellent wind resistance on account of positive attachment.

Mechanically Fastened. The substrate, ie, insulation, cover board, etc, which the single-ply membrane must go over, is typically mechanically fastened to the deck using a low density of fasteners. However, there are also applications where the membrane is applied directly to the deck. The typical

method of installing a mechanically fastened roof membrane is to lay the membrane over the substrate and fasten it to the deck using fasteners spaced along only one of the machine edges of the sheet. The next sheet is then seamed over the fastener to the first sheet, making the system water-tight with the second sheet's loose edge fastened to the deck. This process continues across the roof deck. The maximum slope normally done with a mechanically fastened roof system is 150 cm m (18 in. in a foot). For slopes greater than this, special provisions must be made.

There are also mechanically fastened systems where large sheets are laid over the substrate and seamed together. The mechanical fastening system to hold the membrane to the deck is placed at the appropriate density either over the membrane to fasten the system to the deck or under the membrane to which it is affixed.

Mechanically fastening the membrane leaves the membrane loose between the attachment points. This can cause the membrane to billow in strong winds, giving the roof a quilted pillow look. There is nothing functionally wrong with the system, however, there are ways to minimize the billowing by using air barriers, leaving the old roof in place, or using vents. The manufacturer should be consulted as to the best method to control this quilting effect if there is a concern on this factor. As with the fully adhered systems, the mechanically fastened systems offer a smooth surface and excellent wind resistance.

Manufacturing Methods. The single-ply membrane manufacturing process begins with the mixing of the formulation for the sheet. This is where the antioxidants, antiozonants, stabilizers, fire retardants, fillers, reinforcing fillers, plasticizers, and the polymer are mixed together using either an internal batch mixer, where all the ingredients are loaded into a mixer and then discharged when mixed, or through a continuous mixer, where the ingredients are added at specific locations along the mixer, typically an extruder, as the polymer moves through it. The mixed formulation is then processed generally through either a calender consisting of large opposing rolls through which the material passes, or an extruder that forces the material through a die to generate the flat sheet. Two sheets are usually made and laminated together, whether the construction is reinforced or nonreinforced for extra security against any leak. If a hole is made by accident in one sheet, the chances of a hole in the two laminates lining up is basically zero. There are processes where the reinforced sheet is spread or dip-coated to build up the sheet with the polymer coating.

At this point in the process, thermoplastic and chlorosulfonated polyethylene (CSPE) membranes are complete and are ready for packaging. In the case of ethylene-propylene-diene monomer (EPDM), the curing step occurs before the membrane is ready for packaging. The curing process is accomplished by placing the membrane in a large vulcanizer where the material is heated under pressure to complete the cure.

The flexible single-ply membranes are manufactured in three forms: reinforced, nonreinforced, and fleece-back sheet.

Reinforced. This is a membrane that has been strengthened by the incorporation of a woven or knitted scrim. This reinforcement is typically a 9-by-9 ends per inch knit or a 10-by-10 ends per inch woven 1000 denier (111 tex) polyester, but other materials, such as fiber glass, can be used. The prime use of reinforced membranes is in mechanically fastened roofing systems where the restriction of the elongation of the polymeric material is important for limiting the billowing of the membrane resulting from wind forces. The membrane is used in other installations, ie, fully adhered and ballasted, when improved puncture resistance is sought. In some applications the reinforcement is part of the manufacturing process as a carrier for the polymer. With some polymeric materials, the reinforcement acts as a stabilizer for the sheet to reduce stresses that can occur on the roof.

Nonreinforced. Many nonreinforced elastomeric membranes are available. These allow the roofing system to take full advantage of the elongation, which can be as much as 600%, of the elastomeric membrane. This works to its fullest advantage in handling building movement as well as thermal and other stresses. The nonreinforced membranes are most popular in ballasted roofing systems.

Fleece-Back Sheet. A fleece-back sheet is a nonreinforced polymeric membrane that has had a nonwoven mat made of polyester, weighing 101.7–203.4 g m², laminated to the back of the sheet. The prime use of the fleece-back sheet is in the fully adhered roofing systems. The fleece provides the chemical separator, which eliminates the need for an adhesive that is compatible with the specific membrane or a compatible substrate.

Polymeric Materials. The single-ply membranes are made from a wide variety of polymers. The following is a brief description of those polymers and their characteristics. There are three thermosetting-type elastomeric membranes as of this writing (1996): neoprene, CSPE, and EPDM. Neoprene is still used where oil resistance is needed. For instance, Hydrotech uses neoprene flashings, the base of which is hot-set in rubberized asphalt (see ELASTOMERS, SYNTHETIC-POLYCHLOROPRENE).

CSPE. Chlorosulfonated polyethylene (CSPE), a synthetic rubber manufactured by DuPont, is marketed under the name Hypalon. It can be produced as a self-curing elastomer designed to cure on the roof. The membrane is typically reinforced with polyester and is available in finished thicknesses of 0.75 to 1.5 mm. Because CSPE exhibits thermoplastic characteristics before it cures, it offers heat-weldable seams. After exposure on the roof, the membrane cures offering the toughness and mechanical set of a thermoset. The normal shelf life of the membrane for maintaining this thermoplastic characteristic is approximately six months. After the membrane is fully cured in the field, conventional adhesives are needed to make repairs.

The prime installation method is mechanically fastened but fully adhered and ballasted applications can also be used. CSPE exhibits strong resistance not only to weathering but also to a broad range of chemicals and pollutants; it is also inherently ozone-resistant. It can be produced in many colors and the sheet widths are typically 5–6.5 ft (1.5–1.65 m). The physical characteristics of a CSPE sheet have been described (17) (see ELASTOMERS, SYNTHETIC-CHLOROSULFONATED POLYETHYLENE).

EPDM. Ethylene-propylene-diene monomer (EPDM) rubber, the largest segment of the single-ply elastomeric roofing membrane category, comprising 30% of the total low slope roofing market, is an elastomeric compound synthesized from ethylene, propylene, and a small amount of diene monomer. It is generally used in roofing as a cured material, but there are also formulas for products such as EPDM flashing that are designed to cure on the roof. The membrane is typically nonreinforced, allowing for full utilization of its elongation characteristics. However, reinforced membranes are also offered, which are primarily used in mechanically fastened roofing systems. The membrane thicknesses are typically 1.1 and 1.5 mm, but can also run from 1 to 2.2 mm. Because EPDM exhibits thermoset characteristics, the seaming together of two sheets is accomplished by adhesive systems. Adhesive systems consist of cleaners, cleaner primers, and primers used in various combinations with butyl liquid or tape adhesives. Sealants (qv) are added to some adhesive systems for protection and redundancy (see ELASTOMERS, SYNTHETIC-ETHYLENE-PROPYLENE-DIENE RUBBER).

EPDM is by far the most widely used material in the ballasted roofing system construction. Because of EPDM's flexibility, very large sheets of up to 10,000 square feet (929 m²) can be delivered to the job site in compact rolls that offer reduced labor on the roof in the seaming process. The typical EPDM sheet size used in ballasted systems is 12 by 30 m and 1.1 mm thick. EPDM is also widely used in both the fully adhered and mechanically fastened roofing systems. In these constructions, both 1.1- and 1.5-mm thick material is used with widths from 2.1 to 15 m. A majority of the installations use nonreinforced sheet, although reinforced membrane can also be used in all of the system types. The majority of the reinforced sheets go into mechanically fastened systems.

EPDM membranes are highly resistant to weathering, abrasion, and ozone, as well as a broad range of chemicals. EPDM has excellent low temperature flexibility and excellent resistance to acids, alkalies, and oxygenated solvents such as ketones, esters, and alcohol. However, exposure to aromatic, halogenated, and aliphatic solvents as well as animal fats and vegetable oils should be avoided to prevent swelling and distortion of the membrane. The membrane is offered primarily in black because of superior weathering, but white is also available. The physical characteristics of an EPDM sheet have been described (16).

Thermoplastics. There are five elastomeric membranes that are thermoplastic. Two materials, chlorinated polyethylene (CPE) and polyisobutylene (PIB), are relatively obscure. Thermoplastic materials can be either heat-fused or solvent-welded. In contrast to Hypalon and uncured EPDM, this ability to fuse the membranes together remains throughout the life of the material. However, cleaning of the membrane surface after exposure to weather is required. Correct cleaning procedures for specific membranes are available from the individual manufacturer.

CPA. Copolymer alloy membranes (CPAs) are made by alloying high molecular weight polymerics, plasticizers, special stabilizers, biocides, and antioxidants with poly(vinyl chloride) (PVC). The membrane is typically reinforced with polyester and comes in finished thicknesses of 0.75–1.5 mm and widths of 1.5–1.8 m. The primary installation method is mechanically fastened, but some fully adhered systems are also possible. The CPA membranes can exhibit long-term flexibility by alleviating migration of the polymeric plasticizers, and are chemically resistant and compatible with many oils and greases, animal fats, asphalt, and coal-tar pitch. The physical characteristics of a CPA membrane have been described (15).

EIP. These membranes are compounds consisting of ethylene interpolymers (EIP), PVC, stabilizers, pigments (qv), antioxidants, and modifying polymers. Generally reinforced with polyester fabric, EIP membranes are usually 0.81 mm thick and come in sheet widths of 3.05–6.1 m. The main installation method is mechanically fastened. EIP membranes possess good resistance to chemicals and oils and have high tear strength. Many formulations utilizing combinations of ethylene polymers with other basic ingredients may be produced. The sheet is usually white in color. As of this

writing (1996), an ASTM committee is working to develop a standard for the EIP membranes.

NBP. These nitrile alloy membranes are compounded from PVC, flexibilized by the addition of butadiene–acrylonitrile copolymers, PVC, and other proprietary ingredients. Typically reinforced with polyester scrim, NBP membranes are 1 mm thick and have a width of 1.5 m. They are predominantly used in mechanically fastened roofing systems. NBP membranes exhibit excellent tear and puncture resistance as well as good weatherability, and remain flexible at low temperatures. They are resistant to most chemicals but are sensitive to aromatic hydrocarbons. The sheet is usually offered in light colors. The physical characteristics of NBP membranes have been described (15).

PVC. Poly(vinyl chloride) (PVC), a very versatile polymer, is manufactured by the polymerization of vinyl chloride monomer, a gaseous substance obtained from the reaction of ethylene with oxygen and hydrochloric acid. In its most basic form, the resin is a relatively hard material that requires the addition of other compounds, commonly plasticizers and stabilizers as well as certain other ingredients, to produce the desired physical properties for roofing use. The membranes come in both reinforced and nonreinforced constructions, but since the 1980s the direction has been toward offering only reinforced membranes. The membrane thickness typically runs from 0.8–1.5 mm and widths typically in the range of 1.5–4.6 m.

The greater portion of PVC is installed in the mechanically fastened roofing system; a lesser portion is installed in fully adhered applications. Although PVC was once heavily used in ballasted roofing systems, there are only a small number installed in the 1990s. Fleece-back membrane is popular in the PVC construction for both fully adhered applications as well as in applications where a separator sheet is needed. PVCs are resistant to various weather conditions, bacterial growth, and industrial chemicals. These membranes are chemically incompatible with bituminous materials. PVCs are offered in a variety of colors. The physical characteristics of a PVC membrane have been described (15).

TPO. Thermoplastic polyolefin (TPO) is made by blending ethylene– propylene polymers with polypropylene. The original method of blending the two polymers was by conventional mixing. A more recent technology is to make the two polymers, ethylene–propylene and polypropylene, in the chemical reactor simultaneously, creating a very homogeneous mixture that offers unique properties compared with the original method. The polymer is formulated with stabilizers, pigments, and antioxidants to obtain the appropriate weathering and physical properties for roofing applications. Because of the polymer's inherent flexibility, no plasticizer is required to obtain a flexible sheet. The sheet comes in both reinforced and nonreinforced constructions. Its thickness ranges from 1.1–1.5 mm; its width runs from 1.8–3.0 m.

TPO membrane is most often used in mechanically fastened roofing systems. The reinforced construction is the preferred membrane. Fully adhered systems are installed using both reinforced and nonreinforced constructions. Some ballasted applications are also utilized which use the nonreinforced membrane. TPO membranes are highly resistant not only to a broad range of chemicals, but also weathering, abrasion, and ozone. In addition, they have good low temperature flexibility and excellent resistance to acids, alkalies, and oxygenated solvents such as ketones, esters, and alcohol. However, exposure to aromatic, halogenated, and aliphatic solvents as well as animal fats and vegetable oils should be avoided to prevent swelling and distortion of the membrane. The membrane is offered in a variety of colors. As of this writing, an ASTM committee is working to develop a standard for the TPO membranes.

Modified Bitumen Membrane Systems. The third main type of roofing system available is a hybrid between the conventional built-up roofing products and the elastomeric single-membrane materials (26). Modified bituminous membranes were developed in Europe in the mid-1960s and have been in use in the United States since 1975. These products are made by adding a polymer to an asphalt, which then raises the softening point of the mixture, along with fillers and other processing ingredients, to form a waterproofing sheet. These products are primarily reinforced with polyester or fiber glass nonwoven mats. Some products can have both fiber glass and polyester mats to form a more specialized modified bitumen product. Two basic polymers are used to manufacture modified bitumen products: atactic polypropylene (APP) and styrene–butadiene–styrene (SBS) block copolymers.

APP. Atactic polypropylene is a low crystallinity (amorphous) thermoplastic material used in bitumen modification. The thermoplastic nature of the modifier gives the product excellent rheological properties for heat welding with a torch, and increases the softening point from 32.2°C (90°F) to 148.9°C (300°F). APP not only imparts good weatherability to the asphalt blend, but provides greater toughness to the roof membrane. One adverse feature of APP, which softens at temperatures above 148.9°C (300°F), is its brittleness at temperatures below 6.7°C (20°F). This is a concern on buildings in cold climates that experience abnormal amounts of movement or impact from tools or other falling objects.

SBS. The other widely used modifier in asphalt is a block copolymer, styrene–butadiene–styrene. SBS is a thermoplastic rubber that has rheological properties different from those of APP. SBS addition to an asphalt increases the flexibility over a wider temperature range than does APP. SBS-modified asphalt has excellent elongation properties, with reversibility, and remains flexible at temperatures below 23.3°C (-10°F). The addition of SBS to an asphalt can also increase the softening point from 32.2°C (90°F) to 137.8°C (280°F). The modified bitumen products are tested according to ASTM D5147 (32). There is activity in Asphalt Roofing Manufacturers Association (ARMA) and in ASTM to develop standards for both APP and SBS roofing membranes. However, as of 1996, none exist.

Steep Roofing Products

Asphalt roofing shingles and related products are classified under three broad groups: shingles, roll roofing, and underlayment (34). Shingles and roll roofing are outer roof coverings, ie, they are exposed to the weather and are designed to withstand the elements. Underlayments are inner roof coverings that provide the necessary protection beneath the exposed roofing materials. Asphalt shingles and roll roofing contain two basic components: (1) a base material made of an organic felt or fiber glass mat which serves as the matrix that supports the other components and gives the product the strength to withstand manufacturing, handling, installation, and service conditions; and (2) a specially formulated asphalt coating that provides the long-term weatherability and stability under service temperature extremes.

Shingles. Asphalt shingles (33) are the most common roofing material used in the United States in the 1990s, amounting to more than 80% of all residential roofing products. They are manufactured as strip shingles, laminated (multithickness) shingles, interlocking shingles, and large individual shingles in a variety of weights and colors.

Strip shingles are typically rectangular and may have as many as five cut-outs along the long dimension. Cut-outs separate the shingle tabs, which are exposed to the weather, and give the roof the appearance of being comprised of a larger number of individual units. Strip shingles can also be manufactured without cut-outs to produce a much different appearance. The three-tab shingle is the most common type of strip shingle.

Most of the shingles are available with strips or spots of a factory-applied, self-sealing adhesive, which is a thermoplastic material activated by heat from the sun after the shingle is on the roof. Exposure to the sun's heat bonds each shingle securely to the one below for greater wind resistance. This self-sealing action varies, depending on the geographic location, roof slope, season of the year, and solar orientation.

Weather-resistant mineral granules applied to the top surface of strip shingles during the manufacturing process make possible a wide range of colors. Sand, talc, or mica is applied to the back surface.

The tabs of a strip shingle may be cut either straight or offset to obtain a straight or staggered buttline, respectively. They also may be embossed or built up from a number of laminates of base material to give a three-dimensional effect. Each of these shingle characteristics, ie, staggered buttlines, embossing, and laminate, can be combined in various ways to create textures on the finished roof surface that resemble tile, wood, or slate.

Interlocking shingles are also available and are designed to provide immediate resistance to strong winds. These shingles come in various shapes and have various types of locking designs that provide a mechanical interlock on the roof. Large individual shingles are generally rectangular or hexagonal in shape. Table 7 lists ASTM tests and standards for asphalt shingles.

Table 7. ASTM Tests and Standards for Asphalt Shingles

ASTM designation	Title
D225 ^a	Standard Specification for Asphalt Shingles (Organic Felt) Surfaced with Mineral Granules
D3018 ^b	Standard Specification for Class A Asphalt Shingles Surfaced with Mineral Granules

D3462 ^c	Standard Specification for Asphalt Shingles Made from Glass Felt and Surfaced with Mineral Granules
D3161 ^d	Standard Test Method for Wind Resistance of Asphalt Shingles (Fan-Induced Method)
D228 ^e	Standard Test Methods for Asphalt Roll Roofing, Cap Sheets, and Shingles

^a Asphalt roofing in shingle form; single or multiple thicknesses of organic felt saturated and coated on both sides with asphalt and surfaced on weather side with mineral granules; intended to be applied with exposure to weather <143 mm (5 5 8 in.), and with headlap >51 mm (2 in.). Classified as Type I, uniform-, or nonuniform-thickness shingles of any style; Type II, thick butt, square tab, strip shingles; and Type III, uniform- or nonuniform-thickness shingles of any style.

^b Granule-surfaced asphalt roofing shingles meeting Class A fire test exposure conditions (Test Methods E108); intended to be applied with headlap >51 mm (2 in.). Values (metric) regarded as standard; values in parentheses for information purposes only. Asphalt shingles are Type I, self-sealing; and Type II, nonself-healing.

^c Asphalt roofing in shingle form, composed of glass felt or felts impregnated and coated on both sides with asphalt, and surfaced on the weather side with mineral granules; intended to be applied with headlap >51 mm (2 in.). Must be supplied with a factory-applied self-sealing adhesive or be designed to lock together during installation.

^d Procedures for testing asphalt shingles resistant to wind blowup blowoff when applied on low slopes in accordance with manufacturer instructions. Shingles are Type I, factory-applied adhesive (self-sealing shingles); and Type II, lock-type, with mechanically interlocking tabs (ears).

^e Procedures for physical testing and analyses of roofing and shingles composed of asphalt-saturated or glass fiber felt coated to various extents with asphalt and having coated portion surfaced with powders, laminates, or granules; divided into Type I, single thickness of glass felt coated with asphalt and mineral surfacing (eg, Specifications D2178, D3462, D3672, D3909, and D4601); Type II, single thickness of asphalt-saturated felt coated with asphalt and mineral surfacing (eg, Specifications D224, D225, D249, D2626, and D3672); Type III, similar to Type II, but asphalt-coated and surfaced with mineral granules for part of one side of saturated felt (eg, Specification D371); and Type IV, asphalt roll or shingle roofing formed by laminating >2 mats, felts, papers, foils, fabrics, or films, as web to transport product (eg, Specification D3018).

Roll Roofing. As the name implies, roll roofing is manufactured, packaged, and shipped in rolls. It comes in a wide range of weights and measures. Roll roofing products are produced by applying an asphalt coating to the reinforcing core, and finished with either a smooth or mineral granule surface. Some mineral-surface roll roofing are manufactured with a granule-free selvage edge that indicates the amount each succeeding course should overlap the preceding course. The manufacturer's recommendations with respect to the top, side, and end laps should be followed. The amount of overlap determines how much of the material is exposed to the weather and the extent of coverage to the roof surface; ie, whether most of the surface has a single or a double layer of roll roofing. In addition to its use as a roof covering, roll roofing is also important as a flashing material. Table 8 lists ASTM designations for roll roofing.

Table 8. ASTM Tests and Standards for Roll Roofing

ASTM Designation	Title
D224 ^a	Standard Specification for Smooth-Surfaced Asphalt Roll Roofing (Organic Felt)
D249 ^b	Standard Specification for Asphalt Roll Roofing (Organic Felt) Surfaced with Mineral Granules
D371 ^c	Standard Specification for Asphalt Roll Roofing (Organic Felt) Surfaced with Mineral Granules (Wide Selvage)

^a Sheet form, composed of organic roofing felt, saturated with asphalt and coating on both sides with asphalt compound that may or may not contain mineral stabilizer, surfaced with powdered talc, mica, or other fine mineral matter to prevent sticking. Classified, in mineral net mass per unit area of roofing, as Type I, 1943 g m² (39.8 lb 100 ft²); Type II, 2666 g m² (54.6 lb 100 ft²); Type III, 2495 g m² (51.1 lb 100 ft²); and Type IV, 1943 g m² (39.8 lb 100 ft²).

^b Sheet form, in widths as agreed upon by purchaser and seller, composed of asphalt-saturated organic felt coated on both sides with asphalt and surfaced on weather side with mineral granules, except for selvage. Classified, in minimum net mass of granule-surfaced portion, as Type I, 3610 gm m² (74.0 lb 100 ft²); and Type II, 3490 gm m² (71.5 lb 100 ft²).

^c Sheet form, 914 mm (36 in.) in width, or widths agreed upon by purchaser and supplier, composed of asphalt-saturated organic felt with approximately half the width of weather side coated with asphalt and surfaced with mineral granules, for use as cap sheet in construction of BUR. Materials covered by this specification, in minimum mass per unit area, are Type I, 1806 g m² (37.0 lb 100 ft²); Type II, 2260 g m² (46.3 lb 100 ft²); Type III, 1733 g m² (35.5 lb 100 ft²); and Type IV, 2090 g m² (42.8 lb 100 ft²).

Saturated Felts and Underlayments. These products consist of a dry felt which may be impregnated or coated with an asphalt saturant. They are used primarily as underlayment for asphalt shingles, roll roofing, and other types of roofing materials. These types of felts are also useful as sheathing paper. Saturated felts are manufactured in a variety of weights for use as underlayment or heavy-duty underlayment. Table 9 lists ASTM designations for saturated felts.

Table 9. ASTM Tests and Standards for Saturated Felts

ASTM Designation	Title
D226 ^a	Standard Specification for Asphalt-Saturated Organic Felt Used in Roofing and Waterproofing
D4869 ^b	Standard Specification for Asphalt-Saturated Organic Felt Shingle Underlayment Used in Roofing

^a With or without perforations, may be used with asphalts conforming to Specification D312 requirements in construction of BUR, and Specification D449 requirements in membrane system of waterproofing. Felts covered by this specification are Type I (No. 15 asphalt felt); and Type II (No. 30 asphalt felt).

^b Shingle as underlayment with asphalt shingles. Felts covered by this specification are Type I, shingle underlayment; and Type II, heavy-duty shingle underlayment.

Specialty Eaves Flashing Membranes. Specialty eaves flashing membranes are polymer-modified bituminous sheet materials that are used as underlayment to resist leakage of shingle roofs as a result of ice dams. These sheet materials contain an adhesive layer which is exposed by removing its protective sheet. Special eaves flashing membranes can also be useful in warmer climates where a similar backup of water can occur from an accumulation of pine needles and leaves.

Fire Resistance. Asphalt roofing shingles and rolls are tested by independent laboratories in accordance with established fire-resistance standards. The most widely accepted standard for fire resistance in building materials is ASTM E108, Standard Test Methods for Fire Tests of Roof Coverings. If the material meets the standard, the product may carry the testing laboratory's label indicating its class of fire resistance in accordance with the named standard. These classes are Class A, severe exposure to fire; Class B, moderate exposure to fire; and Class C, light exposure to fire. Most glass

fiber-based asphalt roofing shingle products manufactured in the United States carry a Class A rating, although many organic felt shingles carry the Class C rating.

Wind Resistance. Asphalt shingles are certified to wind performance test standards on a continuous basis through independent third-party testing laboratories. Shingles that have passed the standard wind performance requirements, such as ASTM D3161, Standard Test Method for Wind Resistance for Asphalt Shingles, are identified by labels from the testing laboratory with whom they are in compliance.

Other Shingle Products. Making up about 20% of the steep market are wood shingles and shakes, concrete and clay tiles, natural and mineral fiber slates, and various styles of metal products (34).

Tile. A versatile product, tile is traced back to ancient Greece, China, and Japan. Tile (35) has high fire resistance and is offered in a variety of textures and styles. Both clay and concrete tiles are relatively heavy and require a more robust structure than those used for asphalt shingles.

Slate. Used for hundreds of years, slate comes from different geographical areas and varies in color and weight. Slate (36) is easily split to size and shape; the most common is 4.76-cm (3 16-in.) thick and can weigh 31.71 kg m² (650 lb 100 ft²) or more. Imitation products of mineral fiber and cement stimulate its durability and style.

Wood Shingles and Shakes. Early roofs in the United States were primarily hand-split hickory or cypress shakes. The natural beauty and style of these materials make them popular. Fire-retardant treatment and underlays may be needed to meet local fire codes. Wood shingles are sawn cedar having a uniform thickness. Wood shakes are usually hand-split and resawn.

Metal. Metal shingles and stamped panels that imitate wood or tile are available. These are light in weight and come in many colors and styles, from pre-engineered to custom-fabricated.

Roofing System Performance

Performance (effectiveness) of a roof covering is determined not only by environmental conditions, but also by the individual components that make up the roofing system. The design should be such that no component is underdesigned, compromising the overall performance. Each component should be durable to contribute to the overall roof performance. Procedures for accelerated testing of an entire roofing system are still in the developmental stage as of the mid-1990s. Most criteria are based on durability tests of components, determination of system design, and the proper combination of the roof elements. Roof coverings are subjected to external abuse, eg, foot or other traffic, in addition to weather exposure. The majority of roof leaks are attributed to the defects at penetrations or flashings, so care must be taken to design and install flashings properly.

Weathering. Bitumen hardens progressively owing to ultraviolet attack, oxidation reactions, and, to some extent, loss of plasticizing oils, especially with coal tars. Temperature changes cause expansion and contraction of the roof covering. Other weather factors include moisture, ice, and hail. Wind affects shingles as well as BUR membrane performance. In an accelerated weather study of 15 coating-grade asphalts, a sixfold variation in durability was reported. Performance is correlated to properties and composition. Low asphaltene and high resin contents are desirable. As weathering progresses, the brittle-point temperature, taken by the Fraass procedure, increases. Blending of fractions or base stocks from different crudes improves performance more than antioxidants.

In a comprehensive review of theoretical and practical aspects of asphalt durability, six weathering factors were considered. A test apparatus for the accelerated weathering of organic materials and procedures was developed at the U.S. National Institute of Standards and Technology (formerly the National Bureau of Standards); a carbon arc was used as the energy source. However, outdoor weather exposures are standard procedures for prepared roofing. The opaque granules and mineral stabilizers in the coating asphalts offer excellent protection against the destructive effects of radiant energy and, even in the accelerated Weather-o-meter, many daily cycles are required for failure. Finely divided mineral stabilizers affect the weatherability of coating-grade roofing asphalts. Additions up to 60 wt % of mineral additives can increase durability. Platy minerals such as ground slates and mica (qv) are the most effective. The nature of the asphalt, however, is the most important consideration.

A number of laboratory procedures even more rapid than the accelerated Weather-o-meter have been described for the determination of expected weatherability of coating asphalts. Research sponsored by the Asphalt Roofing Manufacturers Association describes a stepwise procedure to determine changes in the crude asphalt source (see ASPHALT).

Thermal Effects. Temperature influences the oxidation rate, whereas temperature changes impart a mechanical stress to roof coverings. Daily temperature changes, sometimes quite abrupt, can result in a fatigue mechanism causing tensile failure. Repeated cycles of straining, especially at a high level, diminish the strength of roofing felts to the rupture point. Strain reduction by partial fixing of the membrane to the deck offers the biggest improvement in performance.

Data for thermal movement of various bitumens and felts and for composite membranes have been given (1). These describe the development of a thermal shock factor based on strength factors and the linear thermal expansion coefficient. Tensile and flexural fatigue tests on roofing membranes were taken at 21 and 18°C, and performance criteria were recommended. A study of four types of fluid-applied roofing membranes under cyclic conditions showed that they could not withstand movements of <1.0 mm over joints. The limitations of present test methods for new roofing materials, such as prefabricated polymeric and elastomeric sheets and liquid-applied membranes, have also been described (1). For evaluation, both laboratory and field work are needed.

Water Effects. Water in its different forms, ie, liquid, vapor, hail, and ice, profoundly influences the performance of roof coverings. Moisture migrations through roof insulation, eg, vapor that accumulates and later liquefies or water that leaks through the roof covering, reduces insulating efficiency and leads to physical deterioration of the roofing material. Moisture can be detected by direct cut tests, electrical-resistance moisture meters, ir scanning techniques, or nuclear moisture meters.

Flashing flaws in some roofs can be directly responsible for water leakage. The National Roof Contractors Association (NRCA) has developed criteria for curb and edge detail. The recommended flashings include those for stacks, pipes, and expansion joints. Factory Mutual Systems (FM) prescribe perimeter-flashing details dealing primarily with wind-uplift forces that can be encountered on built-up roofs. Roof edge damage in windstorms can be extensive and may result in loss of the roofing membrane, which leads to rain leakage and interior damage.

In roof coverings, organic felts are highly susceptible to expansion and contraction movements with moisture change. Shingles, with a low degree of asphalt saturation, especially if the felt has not been uniformly saturated, show warping. This is described as fish mouthing or clawing. Erratic movements in built-up roofing can be produced by moisture–thermal effects and may cause ridging and failure of the membrane by cracking.

The absorbed equilibrium moisture content of built-up roofing membranes varies with changing environmental conditions. In faulty designs, moisture accumulates during late fall and winter in amounts exceeding those that the system can accommodate in the summer. The weakening effects of moisture are substantially less on fiber glass membranes that possess low equilibrium moisture contents. Organic felt laminates show more movement owing to humidity changes than to temperature changes. Glass felts show small and moderate dimensional changes in response to humidity and temperature changes, respectively; these changes are essentially nondirectional. Roof coverings should be able to withstand 1S in. diameter hailstones without damage. Icing is a function of the roof slope, roof covering composition, and amount of sun exposure.

Fire and Wind Hazards. Weather resistance of roof coverings is not necessarily correlated to fire and wind resistance. Underwriters' Laboratory and the Factory Mutual System test and rate fire and wind hazard resistance, and some durability tests. Organic felt or fiber glass mat base shingles are commonly manufactured to meet minimum UL requirements, which, in addition to minimum mass, require wind and fire resistance properties.

Fire Ratings. Above-deck fire hazards are rated by following ASTM E108 for propagation of the flame along the surface of the roof covering and on the penetration of the fire into the deck or structure. Surface-burning characteristics are measured by either the spread-of-flame test, for noncombustible decks, or the burning brand and intermittent fire tests for combustible decks. The test deck, with a roofing system applied (either a shingle, BUR, or single-membrane construction), is set at a specific incline or slope and exposed to a standard gas flame or burning brand. A list of classified systems is published by each testing agency.

Metal deck assemblies are tested by UL for under-deck fire hazard by using their steiner tunnel (ASTM E84). The assembly, exposed to an under-deck gas flame, must not allow rapid propagation of the fire down the length of the tunnel. FM uses a calorimeter fire-test chamber to evaluate the hazard of an under-deck fire. The deck is exposed to a gas flame and the rate of heat release is measured and correlated to the rate of flame propagation. A different FM test assesses the damage to roof insulations exposed to radiant heat.

Wind Testing. UL employs two wind resistance tests, one for shingles and another for BUR assemblies. ASTM D3161 is a standard procedure for measuring the wind resistance of asphalt shingles. In ASTM D3161 and the UL shingle test, the conditioned test deck is placed at a specified slope and exposed to a wind velocity of 97 km/h (60 mph) for two hours or until either failure by tab or shingle lifting is known.

Factory Mutual provides loss prevention data sheets that explain how to protect buildings from wind damage. Pressure coefficients that define increased uplift at corners and edges adjust the calculated uplift pressures. A laboratory uplift pressure test rates roofing assemblies. An uplift pressure of 2.9 kPa (0.42 psi) must be withstood under FM conditions to meet the Class 1-60 requirements. The FM approval guide is revised annually (37).

Economic Aspects

The National Roofing Contractor's Association conducted a survey of U.S. contractors' total sales, the type of materials used, the percentage of low slope work performed, and the amount contributed by new construction and reroofing in 1995 (38). In 1995, low slope roofing was estimated to take up 71.9% of the total roofing market, at \$13.2 billion, of which 74.9% was dominated by the reroofing segment. According to the survey, the systems most often specified for low slope reroofing and new construction were single plies, eg, EPDM, CSPE, and PVC. These membranes captured 32.7% of the reroofing market and 40.5% of the new construction market. The built-up roofing systems had 26.8% of the reroofing sector and 23.5% of the new construction market. Modified bitumen roof membrane had 23.2% of the reroofing segment and 16.9% of the new construction market. APP and SBS modifieds were equally employed. Sprayed polyurethane foam (SPF) usage for the reroofing and new construction markets was 2.6% and 2.3%, respectively. Liquid-applied systems had 2.2% in reroofing and 0.9% in new construction.

The steep-slope roofing market amounted to 28.1% of the total roofing market, at \$5.15 billion in 1995. In the steep-roofing market, 81.3% of the business was reroofing, and new construction accounted for 18.7%. As for reroofing materials, asphalt shingles, which include fiber glass (54.3%) and organic (13.7%), continue to dominate the steep-slope roofing market with a 68% share. Other materials used in the reroofing segment were low slope materials (19.9%), tile (2.6%), metal (2.8%), wood shingles/shakes (2.4%), slate (2%), and other (2.3%). In the new construction market, estimates are made for the following materials used: asphalt shingles (63.1%), low slope materials (18.1%), tile (6.3%), metal (2.6%), wood shingles/shakes (4%), slate (3.5%), and other (2.4%).

Health, Safety, and Environmental Factors

The materials used by the roofing industry are constantly being scrutinized for health, safety, and environmental requirements. Because the regulations governing the use of these materials change rapidly, it is difficult to indicate where all of the materials stand. A partial listing of the issues that have appeared from 1980–1995 include the Clean Air Act, fiber glass health, asphalt fumes, volatile organic compounds (VOCs), replacement of chlorinated fluorocarbons (CFCs) by hydrochlorinated fluorocarbons (HCFCs), elimination of asbestos products, and greater emphasis on recycling. This list is only a small sampling of the issues that have surrounded the roofing industry. Any study of this industry should always involve a review of the latest regulations regarding the use of any products.

BIBLIOGRAPHY

"Roofing Materials" in *ECT* 2nd ed., Vol. 17, pp. 459–475, by G. W. Berry, Johns-Manville Corp.; in *ECT* 3rd ed., Vol. 20, pp. 320–336, by A. J. Hoiberg and E. G. Long, Manville Corp.

1. ASTM D226, Asphalt-Saturated Organic Felt Used in Roofing and Waterproofing, ASTM, Philadelphia, Pa.
2. ASTM D2626, Asphalt-Saturated-and-Coated Organic Felt Base Sheet Used in Roofing, ASTM, Philadelphia, Pa.
3. ASTM D2178, Asphalt Glass Felt Used in Roofing and Waterproofing, ASTM, Philadelphia, Pa.
4. ASTM D4601, Asphalt-Coated Glass Fiber Base Sheet Used in Roofing, ASTM, Philadelphia, Pa.
5. ASTM D4897, Asphalt-Coated Glass Fiber Venting Base Sheet Used in Roofing, ASTM, Philadelphia, Pa.
6. ASTM D3909, Asphalt Roll Roofing (Glass Felt) Surfaced with Mineral Granules, ASTM, Philadelphia, Pa.
7. ASTM D312, Asphalt Used in Roofing, ASTM, Philadelphia, Pa.
8. ASTM D1863, Mineral Aggregate Used on Built-Up Roofs, ASTM, Philadelphia, Pa.
9. ASTM D2824, Aluminum-Pigmented Asphalt Roof Coatings, Non-Fibered, Asbestos, and Fibered, and Fibered without Asbestos, ASTM, Philadelphia, Pa.
10. ASTM D1227, Emulsified Asphalt Used as a Protective Coating for Roofing, ASTM, Philadelphia, Pa.
11. ASTM D4479, Asphalt Roof Coatings, Asbestos-Free, ASTM, Philadelphia, Pa.
12. ASTM D41, Asphalt Primer Used in Roofing, Dampproofing, and Waterproofing, ASTM, Philadelphia, Pa.
13. ASTM D450, Coal-Tar Pitch Used in Roofing, Dampproofing, and Waterproofing, ASTM, Philadelphia, Pa.
14. ASTM D3019, Lap Cement Used with Asphalt Roll Roofing, Non-Fibered, Asbestos Fibered, and Non-Asbestos Fibered, ASTM, Philadelphia, Pa.
15. ASTM D4434, Standard Specification for Poly(Vinyl Chloride) Sheet Roofing, ASTM, Philadelphia, Pa.
16. ASTM D4637, Standard Specification for Vulcanized Rubber Sheet Used in Single-Ply Roof Membrane, ASTM, Philadelphia, Pa.
17. ASTM D5019, Standard Specification for Reinforced Non-Vulcanized Polymeric Sheet Used in Roofing Membrane, ASTM, Philadelphia, Pa.
18. *Proceedings of the 7th Conference on Roofing Technology*, National Bureau of Standards (NBS), Gaithersburg, Md., and National Roofing Contractors Association (NRCA), Rosemount, Ill., Apr., 14–15, 1983.
19. *Proceedings of the 8th Conference on Roofing Technology*, NBS and NRCA, Apr. 16–17, 1987; see Ref. 18.
20. *Proceedings of the 9th Conference on Roofing Technology*, NBS and NRCA, May, 4–5, 1989; see Ref. 18.
21. *Proceedings of the 10th Conference on Roofing Technology*, NBS and NRCA, Apr. 22–23, 1993; see Ref. 18.
22. *Proceedings of the 11th Conference on Roofing Technology*, NBS and MRCA, Sept. 21–22, 1995; see Ref. 18.
23. *Second International Symposium on Roofing Technology*, International Union of Testing and Research Laboratories for Materials & Structures, NBS, and NRCA, 1985.
24. *Third International Symposium on Roofing Technology*, National Institute of Standards and Technology, Gaithersburg, Md., Canadian Roofing Contractors Association, National Research Council of Canada, International Waterproofing Association, NRCA, and International Council for Building Research Studies and Documentation, 1991.
25. ANSI RP-4, Wind Design Guide for Ballasted Single-Ply Roofing Systems, ANSI, New York.
26. *Flexible Membrane Roofing: A Professional's Guide to Specifications*, SPRI, Needham, Mass.
27. M. J. Rosenfield, *Construction of Experimental Polyvinyl Chloride (PVC) Roofing*, U.S. Army Corps of Engineers, Champaign, Ill., 1984.
28. H. O. Laaly, *The Science and Technology of Traditional and Modern Roofing Systems*, Los Angeles, Calif., 1992.
29. *Commercial Low Slope Roofing Materials Guide*, National Roofing Contractors Association, Rosemont, Ill.
30. *Loss Prevention Data, Adhered or Mechanically Attached Single-Ply Membrane Roof Systems 1–29*, Factory Mutual System, Norwood, Mass.
31. *Loss Prevention Data, Loose-Laid Ballasted Roof Coverings 1–29*, Factory Mutual System, Norwood, Mass.
32. ASTM D5147, Standard Test Methods for Sampling and Testing Modified Bitumen Sheet Materials, ASTM, Philadelphia, Pa.
33. *Residential Asphalt Roofing Manual*, Asphalt Roofing Manufacturers Association, Rockville, Md., 1993.
34. R. H. Herbert III, *Roofing: Design, Criteria, Options Selection*, R. S. Means Co., Inc., Kingston, Mass., 1989.
35. ASTM C1167, Clay Roof Tile, ASTM, Philadelphia, Pa.

36. ASTM C406, Roofing Slate; ASTM C1225, Non-Asbestos Fiber-Cement Shingles; ASTM C1185, Sampling and Testing Non-Asbestos Fiber-Cement; ASTM C121, Test Method for Water Absorption of Slate; ASTM C217, Test Method for Weather Resistance of Natural Slate, ASTM, Philadelphia, Pa.

37. *Factory Mutual Approval Guide*, Factory Mutual System, Norwood, Mass., published annually.

38. *Prof. Roofing* (Mar. 1996).

Ruben G. Garcia
Ray L. Corbin
Schuller International, Inc.

Richard J. Gillenwater
Carlisle Syntec Systems

Jim Compton
Globe Building Materials Company

Cliff Patenaude
Bird, Inc.

Brian Anthony
The Brewer Company

Jim Hunter
Gardner Asphalt Corporation



ROSANILINE.

See [TRIPHENYLMETHANE AND RELATED DYES](#).

ROSIN AND ROSIN DERIVATIVES.

See [RESINS, NATURAL](#); [TERPENOID](#)S.

RUBBER CHEMICALS

Rubber chemicals are materials that are added in minor amounts to rubber formulations in order to improve their properties and make them commercially useful. Raw rubber polymer has very limited practical applications because of tackiness, flow, and other undesirable features. Rubber chemicals are added to assist processing, promote cross-linking, and provide longevity to the part in service. Vulcanizing adjacent polymer chains together by cross-links prevents flow, increases strength, and provides recovery from deformation. The most widely used method of cross-linking polymer chains is by heating with elemental sulfur (vulcanization). Accelerators speed up the reaction of sulfur with polymer to improve the economics of manufacture and prevent degradation that would otherwise occur upon prolonged heating. Peptizers and process aids assist flow during the mixing and shaping operations. Antidegradants protect the part in service from heat, oxygen, ozone, and repeated flexing. Other rubber chemicals function as blowing agents, adhesion promoters, and activators or retarders which modify the onset of cross-linking.

Organic chemicals which are used primarily in the rubber industry contributed \$415 million to the United States economy in 1994 (1). The historical sales of the primary classes of organic rubber-processing chemicals are summarized in Table 1. Accelerators and antidegradants are the two main types of rubber chemicals. The U.S. production of accelerators has been declining, due to longer-lasting radial tires and to fewer domestic producers. The production of antidegradants shows the effect of the adoption of radial tires in the 1970s but has been increasing more recently because of the desire for articles with longer service life. Antidegradants are discussed elsewhere (see [ANTIOXIDANTS](#); [ANTIOZONANTS](#)). A significant component of the "other" category in Table 1 is alkyl mercaptans, used as polymerization regulators. Total organic rubber chemical use has remained stable at about 6% of synthetic rubber production.

Table 1. U.S. Sales of Organic Rubber-Processing Chemicals,^a 10⁶ kg

Chemicals	Year					
	1965	1970	1975	1980	1985	1990
accelerators	31.8	38.7	32.1	29.3	21.3	
benzothiazoles and sulfenamides	20.0	26.5	24.5	21.0	16.4	25.4
dithiocarbamates and thiurams	6.3	7.1	4.5	3.9	2.5	
antidegradants	45.1	52.5	44.5	43.3	47.7	72.9
substituted <i>p</i> -phenylenediamines	13.4	18.6	17.9	15.4	19.8	26.6
other amines	21.6	19.1	18.2	10.9	12.2	20.0
phenols and phosphites	10.1	14.8	8.4	17.0	15.7	26.3
other chemicals	11.0	12.2	15.9	15.4	11.9	
<i>Total</i>	<i>87.9</i>	<i>103.4</i>	<i>92.5</i>	<i>88.0</i>	<i>79.1</i>	<i>136.4</i>

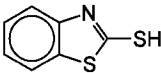
^a Ref. 1.

In addition to the materials shown in Table 1, other organic materials find a minor portion of their use in rubber processing, such as waxes and fatty acids. Also, the rubber industry uses modest amounts of inorganic compounds, notably elemental sulfur, zinc oxide, magnesium oxide, and sodium bicarbonate.

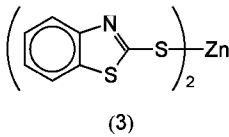
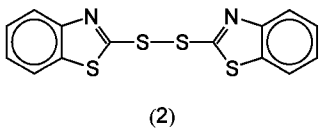
Accelerators of Vulcanization

Vulcanization was first reported in 1839 with the discovery that heating natural rubber with sulfur and basic lead carbonate produced an improvement in physical properties (2). In 1906, aniline was the first organic compound found to have the ability to accelerate the reaction of sulfur with natural rubber (3). Various derivatives of aniline were soon developed which were less toxic and possessed increased acceleration activity.

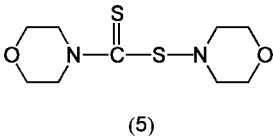
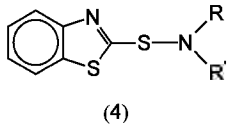
Mercaptobenzothiazoles. These compounds form the basis for the largest-volume organic accelerators. 2-Mercaptobenzothiazole (MBT) (1) is prepared by heating aniline, carbon disulfide, and sulfur in an autoclave at elevated temperature and pressure. Tars from the reaction can be removed by taking the crude MBT up in water by making the sodium salt, clarifying the solution, and then precipitating pure MBT with acid. MBT can be oxidized to the disulfide (MBTS) (2) or it can react with zinc oxide to form the zinc salt (ZMBT) (3). MBTS is the largest-volume of the three benzothiazole accelerators, selling for under \$3 /kg. ZMBT is too fast-curing for most dry rubber compounds and is used mainly in latex applications.



(1)



Sulfenamides. Sulfenamides (4) are often produced by oxidizing an equimolar mixture of MBT and an aliphatic amine. Alternatively, the *N*-chloroamine can react with the sodium salt of MBT. One sulfenamide, OTOS (5), uses a thiocarbamyl functionality in place of the benzothiazole group.



Sulfenamides do not become active accelerators until the sulfur–nitrogen bond thermally dissociates, providing delayed action before cross-linking occurs. Delayed action is useful in applications which require lengthy processing operations such as joining many layers together to make a tire. Sulfenamides are the primary accelerator used in building tires, making them the largest-volume of all accelerator classes. Sulfenamides are priced at over \$4 /kg. Because of environmental concerns of exposure to stable *N*-nitrosamines, tire makers have switched from sulfenamides based on morpholine or other secondary amines to sulfenamides based on primary amines (R' = H).

Dithiocarbamates. These compounds, (6) and (7), are prepared by the reaction of dialkyl amines with carbon disulfide in the presence of sodium hydroxide. For use as rubber accelerators, the sodium salts initially prepared are converted to water-insoluble zinc or other metal salts. Dithiocarbamates provide the most rapid cross-linking and, together with the related thiurams, are known as ultra-accelerators. Various metals are used to achieve specific properties. Copper provides low compression set for O-rings and gaskets based on synthetic rubbers. Selenium and tellurium provide good aging and high modulus. Heavy metals such as cadmium and lead do not migrate and have been used in polymer blends although they are being replaced by bismuth. Sodium and piperidine are exceptionally fast-curing and are used in latex operations. Zinc is most frequently used (see Table 2).

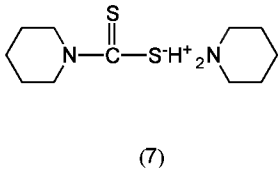
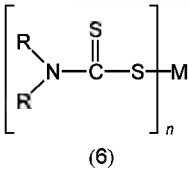


Table 2. Properties of Selected Commercial Accelerators

Name	Acronym	Color	CAS Registry Number	Mol wt	d, mg/cm ³	Melting point, °C	Sol'y, toluene ^a	Acute oral LD ₅₀ , mg/kg ^b
Benzothiazoles								
2-mercaptobenzothiazole (1)	MBT	light yellow–buff	[149-30-4]	167	1.50	175–182	v sol	3,800
bis(2,2'-benzothiazolyl)disulfide (2)	MBTS	cream–light yellow	[120-78-5]	332	1.51	164–179	m sol	7,900
zinc 2-mercaptobenzothiazole (3)	ZMBT	pale yellow	[155-04-4]	398	1.63	300 min	sl sol	>5,000
Sulfenamides								
<i>N</i> - <i>tert</i> -butyl-2-benzothiazolesulfenamide (4, R ¹ = <i>t</i> -C ₄ H ₉)	TBTS	light tan–buff	[95-31-8]	238	1.28	105 min	m sol	>6,300
<i>N</i> -cyclohexyl-2-benzothiazolesulfenamide (4, R = H, R ¹ = C ₆ H ₁₁)	CBTS	cream–light tan	[95-33-0]	264	1.28	94–102	v sol	5,300
<i>N</i> -oxydiethylene-2-benzothiazolesulfenamide (4, RR'N = N(CH ₂ CH ₂) ₂ O)	OBTS	light tan	[102-77-2]	252	1.37	70–90	v sol	>8,200
4-morpholinyl-2-benzothiazole disulfide (4, RR'N = SN(CH ₂ CH ₂) ₂ O)	MBSS	cream–light yellow	[95-32-9]	284	1.51	115–130	sol	11,500
<i>N</i> -dicyclohexyl-2-benzothiazolesulfenamide (4,	DCBS	off-white	[4979-32-2]	34	1.2	90–102	m sol	3,400

R = R' = C ₆ H ₁₁) <i>N-tert</i> -butyl-2-benzothiazolesulfenimide (14)	TBSI	off-white	[3741-80-8]	7 40	0 1.3	128 min		>5,000
<i>N</i> -oxydiethylenethiocarbamyl- <i>N</i> -oxydiethyl-enesulfenamide (5)	OTOS	white–cream	[13752-51-7]	4 24 8	5 1.3 4	136–142	sol	~5,000
Dithiocarbamates								
bismuth dimethyldithiocarbamate (6, R = CH ₃ , n = 3, M = Bi)	BiDMC	lemon yellow	[21260-46-8]	57 0	2.0 4	230 dec	sl sol	>3,000
cadmium diethyldithiocarbamate (6, R = C ₂ H ₅ , n = 2, M = Cd)	CdDEC	white–light gray	[14239-68-0]	40 9	1.3 9	65 min	m sol	7,100 ^c
copper dimethyldithiocarbamate (6, R = CH ₃ , n = 2, M = Cu)	CuDMC	dark brown	[137-29-1]	30 4	1.7 5	325 min	m sol	>5,000
lead diamyldithiocarbamate (6, R = C ₅ H ₁₁ , n = 2, M = Pb)	PbDAC	light amber	[36501-84-5]	67 2	1.1 0	liquid	v sol	>10,000
lead dimethyldithiocarbamate (6, R = CH ₃ , n = 2, M = Pb)	PbDMC	gray	[19010-66-3]	44 8	2.4 3	300 min	p insol	
piperidinium pentamethylenedithiocarbamate		cream–light yellow	[98-77-1]	24 6	1.2 0	163 min	m sol	250
selenium dimethyldithiocarbamate (6, R = CH ₃ , n = 4, M = Se)	SeDMC	yellow	[144-34-3]	56 0	1.5 8	140–172	sol	>200
sodium di- <i>n</i> -butyldithiocarbamate (6, R = n-C ₄ H ₉ , M = Na)	NaDBC	pale amber	[136-30-1]	22 7	1.0 9	liquid	v sol	>16,000
tellurium diethyldithiocarbamate (6, R = C ₂ H ₅ , n = 4, M = Te)	TeDEC	orange-yellow	[20941-65-5]	72 1	1.4 4	108–119	sol	>5,000
Zinc dithiocarbamates (6, M = Zn, n = 2)								
zinc diamyldithiocarbamate (R = C ₅ H ₁₁)	ZDAC	light amber	[15337-18-5]	53 0	0.9 9	liquid	v sol	14,900
zinc dibenzoyldithiocarbamate (R = CH ₂ C ₆ H ₅)	ZBEC	white–light cream	[14726-36-4]	61 0	1.4 2	180–190		>5,000
zinc di- <i>n</i> -butyldithiocarbamate (R = n-C ₄ H ₉)	ZDBC	white–cream	[136-23-2]	47 4	1.2 1	104–112	sol	>16,000
zinc di- <i>iso</i> -butyldithiocarbamate (R = i-C ₄ H ₉)	ZDiBC	white–cream	[36190-62-2]	47 4	1.2 4	110–120	sol	
zinc diethyldithiocarbamate (R = C ₂ H ₅)	ZDEC	white	[14324-55-1]	36 2	1.4 8	171–183	m sol	3,500
zinc dimethyldithiocarbamate (R = CH ₃)	ZDMC	white	[137-30-4]	30 6	1.7 1	242–257	m sol	1,400 ^d
zinc di- <i>iso</i> -nonyldithiocarbamate (R = i-C ₉ H ₁₉)	ZDiNC	off-white		75 4	1.2 4	~77	sol	>5,000
zinc pentamethyldithiocarbamate (R ₂ N = piperidine)	ZPDC	off-white	[13878-54-1]	38 6	1.6 0	225		1,200 (mice)
Thiurams								
dipentamethylenethiuram disulfide (9, x = 2)	DPTD	white–cream	[94-37-1]	32 1	1.4 0	120		>5,000 (mice)
dipentamethylenethiuram hexasulfide (9, x = 6)	DPTH	light yellow–buff	[971-15-3]	44 9	1.5 0	115 min	sol	>200 ^e
dipentamethylenethiuram monosulfide (9, x = 1)	DPTM	yellow	[725-32-6]	28 9	1.4 0	110		
tetrabenzylthiuram disulfide (8, R = CH ₂ C ₆ H ₅ , x = 2)	TBzTD	light yellow	[10591-85-2]	54 5	1.3 1	127–135		>5,000
tetra- <i>n</i> -butylthiuram disulfide (8, R = n-C ₄ H ₉ , x = 2)	TBTD	dark amber	[1634-02-2]	40 9	1.0 6	liquid	v sol	2,300 ^e
tetra- <i>iso</i> -butylthiuram disulfide (8, R = i-C ₄ H ₉ , x = 2)	TiBTD	light yellow	[3064-73-1]	40 9	1.1 4	65–70		
tetraethylthiuram disulfide (8, R = C ₂ H ₅ , x = 2)	TETD	light yellow	[97-77-8]	29 7	1.2 7	67 min	sol	1,300
tetramethylthiuram disulfide (8, R = CH ₃ , x = 2)	TMTD	white–cream	[137-26-8]	24 0	1.4 2	146–156	sol	1,800 ^f
tetramethylthiuram monosulfide (8, R = CH ₃ , x = 1)	TMTM	yellow	[97-74-5]	20 8	1.3 7	103–114	v sol	1,300 ^g
Guanidines								
diphenylguanidine (10, R = C ₆ H ₅)	DPG	white–faint yellow	[102-06-7]	21 0	1.2 0	144–149	sl sol	350
di- <i>ortho</i> -tolylguanidine (10, R = o-C ₆ H ₄ CH ₃)	DOTG	white–pink	[97-39-2]	23 8	1.2 0	173–179	sl sol	134–500
DOTG salt of dicatechol borate	PML	tan–light brown	[16971-82-7]	46 7	1.2 5	165 min	m sol	570
Thioureas								
<i>N,N'</i> -di- <i>n</i> -butylthiourea (11, R = n-C ₄ H ₉)	DBTU	white–light tan	[109-46-6]	18 8	1.0 3	56–65	v sol	350
<i>N,N'</i> -diethylthiourea (11, R = C ₂ H ₅)	DETU	white–light yellow	[105-55-5]	13 2	1.1 1	68–77	m sol	310
<i>N,N'</i> -diphenylthiourea (thiocarbanilide) (11, R = C ₆ H ₅)	DPTU	white–light gray	[102-08-9]	22 8	1.3 0	148	sl sol	1,500
ethylenethiourea (11, R = CH ₂ CH ₂)	ETU	white	[96-45-7]	10 2	1.4 0	195		1,800 ^h
trimethylthiourea (11, R = CH ₃)	TMTU	white	[2489-77-2]	11 8	1.2 3	80–90	m sol	670 ⁱ
Xanthates								
dibutylxanthogen disulfide (12, R = n-butyl, M = Na)	CPB	amber	[105-77-1]	44	1.1	liquid		

not present)				5	5			
dibutylxanthogen polysulfide (12, R = <i>n</i> -butyl, M = S _x)	XPS	yellow			1.2	liquid		1,800
zinc di- <i>iso</i> -propylxanthate (12, R = <i>i</i> -propyl, M = Zn)	ZIX	white–light yellow	[1000-90-4]	33	1.5	140 dec	sl sol	>2,000
		Dithiophosphates		6	6			
copper <i>O,O</i> -di- <i>iso</i> -propylphosphorodithioate (13, R = <i>i</i> -C ₃ H ₇ , M = Cu)	CUT	yellow-green	[41593-12-8]	49	1.1	69		
zinc <i>O,O</i> -di- <i>n</i> -butylphosphorodithioate (13, R = <i>n</i> -C ₄ H ₉ , M = Zn)	ZBPD	yellow-green	[6990-43-8]	54		liquid		
		Aldehyde–amines						
butyraldehyde–aniline reaction			[34562-31-7]		0.9	liquid	v sol	500–5,000
butyraldehyde–monobutylamine			[68411-19-8]		8			
					0.8	liquid	v sol	1,000 (ALD)
					6			
heptaldehyde–aniline reaction			[9003-50-3]		0.9	liquid	v sol	
					3			
hexamethylenetetramine [100-97-0]			[100-97-0]	14	1.3	200 dec		
				0	0			
		Other accelerators						
dimethylammonium hydrogen isophthalate			[71172-17-3]	21	1.3	190 dec	sl sol	>15,000
				1	5			
3-methyl-2-thiazolidinethione			[1908-87-8]	13	1.3	64		1,200
				3	9			
mono- and dibenzyl amine					1.1	liquid	v sol	225
					6			
<i>N</i> -ethylcyclohexyl amine			[5459-93-8]	12	0.8	liquid		
				7	5			
dihydrogenated tallow amine			[31789-79-5]	51	0.8	62		>10,000
				0	0			

^a Solubility range: p insol = <0.1 g solute / 100 mL solution; sl sol = 0.1–<1.0; m sol = 1.0–<5.0; sol = 5.0–<10.0; and v sol = >10.0.

^b Toxicity data are for rats unless otherwise indicated.

^c TDL₀ mice for 78 wks.

^d 320 mg / kg in second study.

^e Intraperitoneal for mice.

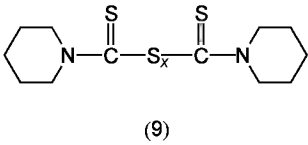
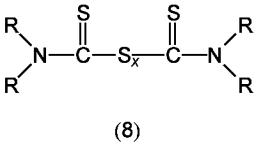
^f For females; 3700 mg / kg males.

^g LD_{LO}: 500 mg / kg rats, 100 mg / kg dogs, 100 mg / kg rabbits, and 10 mg / kg guinea pigs.

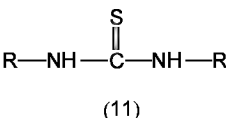
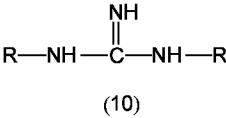
^h Classified as IARC carcinogen.

¹ In structure (11), one hydrogen is replaced with a methyl group.

Thiuram Sulfides. These compounds, (8) and (9), are an important class of accelerator. Thiurams are produced by the oxidation of sodium dithiocarbamates. The di- and polysulfides can donate one or more atoms of sulfur from their molecular structure for vulcanization. The use of these compounds at relatively high levels with little or no elemental sulfur provides articles with improved heat resistance. The short-chain (methyl and ethyl) thiurams and dithiocarbamates are priced ~\$2 / kg. Producers have introduced ultra-accelerators based on longer-chain and branched-chain amines that are less volatile and less toxic. This development is also motivated by a desire to minimize airborne nitrosamines.



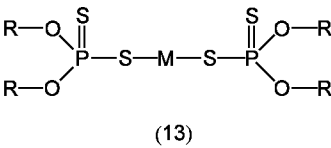
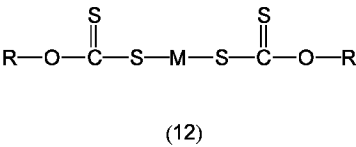
Guanidines. Guanidines (10) were one of the first aniline derivatives used as accelerators. They are formed by reaction of two moles of an aromatic amine with one mole of cyanogen chloride. Diphenylguanidine (DPG) has enjoyed a resurgence in demand as an activator for sulfenamides and a co-accelerator in tire tread compounds which employ silica fillers for low rolling resistance. Guanidines alone show too little activity to be extensively used as primary accelerators. There were no U.S. producers as of mid-1996.



Thioureas. Thioureas (11) are typically made from primary amines and carbon disulfide. The amine can be ethylamine, butylamine,

ethylenediamine, or aniline. The last produces diphenylthiourea which is commonly known as thiocarbanilide. Thioureas are often used as accelerators for polychloroprene (Neoprene) or other rubbers that contain chlorine as reactive cross-linking sites.

Xanthates. These compounds (12) are relatively fast accelerators which are used at low temperature because most examples decompose without cross-linking at higher temperature. Xanthates (qv) are produced by reaction of equimolar amounts of alcohol and carbon disulfide in the presence of caustic. The sodium salt is then converted to the zinc compound or oxidized to the disulfide.



Dithiophosphates. These compounds (13) are made by reaction of an alcohol with phosphorus pentasulfide, then neutralization of the dithiophosphoric acid with a metal oxide. Like xanthates, dithiophosphates contain no nitrogen and do not generate nitrosamines during vulcanization. Dithiophosphates find use as high temperature accelerators for the sulfur vulcanization of ethylene–propylene–diene (EPDM) terpolymers.

Aldehyde–Amines. There is a long history of use of the reaction products of aldehydes and amines. Most such reactions yield a dark-colored liquid mixture containing many different structures. Hexamethylenetetramine (Hexa) is a notable exception, giving an adamantane-type structure with nitrogen atoms replacing carbon at the bridgeheads. Hexa reverts to formaldehyde and ammonia in the presence of water, and is a scorchy accelerator sometimes used in adhesion systems. At one time, aldehyde–amines were used in making hard rubber (ebonite), but this use of rubber has largely been replaced by plastics. Aldehyde–amines find use as accelerators for thiadiazole cures of halogenated polymers.

Other Accelerators. Amine isophthalate and thiazolidine thione, which are used as alternatives to thioureas for cross-linking polychloroprene (Neoprene) and other chlorine-containing polymers, are also used as accelerators. A few free amines are used as accelerators of sulfur vulcanization; these have high molecular weight to minimize volatility and workplace exposure. Several amines and amine salts are used to speed up the dimercapto thiadiazole cure of chlorinated polyethylene and polyacrylates. Phosphonium salts are used as accelerators for the bisphenol cure of fluorocarbon rubbers.

A listing of selected commercial accelerators and their properties is shown in Table 2. Within a family of accelerators, materials of low molecular weight usually provide more cross-links at a given dosage. In general, rubber chemicals with low density are preferred, because rubber chemicals are purchased by the kilogram but rubber articles are sold by the item (volume). Liquid products are often available adsorbed on an inert carrier for easier handling as flowable powders. Solid materials with low melting points (<100°C) can be offered as flakes or pastilles; those products with higher melting points must be ground to fine powders to disperse properly in the mixing operation. Most accelerators dissolve in the hot rubber mixture. Upon standing at room temperature, an accelerator can migrate to the surface (bloom) if its solubility limit is exceeded. Toluene provides a good model to estimate solubility limits in highly unsaturated hydrocarbon rubbers. Accelerators have poorer solubility in predominantly saturated hydrocarbon rubbers such as EPDM and isobutylene–isoprene copolymer (butyl rubber), and greater solubility in polar rubbers such as Neoprene and butadiene–acrylonitrile copolymer (nitrile rubber) (see ELASTOMERS, SYNTHETIC).

Health and Safety Factors. To minimize exposure to potentially toxic materials, powders are generally offered with an antisticking agent applied. Many chemicals, including accelerators, can cause skin irritation in sensitive individuals. The availability of pellets, prills, and polymer-encapsulated master batches to minimize worker exposure to chemicals is a service offered by several suppliers. Accelerators based on secondary amines can liberate small quantities (parts per billion) of *N*-nitrosodialkylamines. Most nitrosamines do not have the five-strain Ames test that is accepted in the 1990s. In some test animals, a few nitrosamines are carcinogenic, with the potency ranking being *N*-nitrosodiethylamine (NDEA) > *N*-nitrosodimethylamine (NDMA) > *N*-nitrosomorpholine (NMOR) > *N*-nitrosonornicotine (NNN) > *N*-nitrosodiethanolamine (NDELA) > *N*-nitrosodiphenylamine (NDPhA) (4). Nitrosamine levels in rubber articles for use by infants are regulated in several countries. Manufacturers of these articles have developed nitrosamine-free compounds. However, great care must be taken in manufacturing and packaging to prevent recontamination of the articles with nitrosamines which are ubiquitous in the environment.

The main producers of organic accelerators for rubber vulcanization are shown in Table 3. This table is not meant to be completely comprehensive, but rather to indicate the principal historical suppliers to the rubber industry. Most producers offer chemical equivalents in the largest-volume products. Within the range of smaller-volume specialty accelerators, chemical equivalents become less common. Each producer may offer different products to achieve the same purpose of rapid cross-linking, resistance to thermal degradation, or other performance characteristics. Many offer proprietary blends of accelerators.

Table 3. Producers of Accelerators

Company	Headquarters location
Bann Quimica, Ltd.	Sao Paulo, Brazil
Bayer AG	Leverkusen, Germany
Chemetal GmbH	Frankfurt-am-Main, Germany
Flexsys, NV	Brussels, Belgium
General Quimica, SA	Burgos, Spain
The B.F. Goodrich Co.	Akron, Ohio
Karbochem	Sloane Park, South Africa
Kawaguchi Chemical Industry Co., Ltd.	Tokyo, Japan
Lestar Quimica, SA	Buenos Aires, Argentina
MLPC International	Rion des Landes, France
Oriental Chemical Industries	Seoul, Korea
Ouchi Shinko Chemical Industrial Co.	Tokyo, Japan
Premier Chemical Co., Ltd.	Taipei, Taiwan
Robinson Brothers, Ltd.	West Midlands, Great Britain
Sanshin Chemical Industry Co., Ltd.	Yamaguchi, Japan
Sumitomo Chemical Co., Ltd.	Tokyo, Japan
UCB, SA	Brussels, Belgium
Uniroyal Chemical Co.	Middlebury, Conn.
R. T. Vanderbilt Co., Inc.	Norwalk, Conn.

Cross-Linking Agents

Sulfur. The most widely used and economical cross-linking agent is sulfur, selling for about \$0.35 /kg. It provides rubber articles with high strength and excellent resistance to failure when flexed. Improved resistance to heat can be obtained by reducing the amount of sulfur and increasing the amount of accelerator. This change provides a greater proportion of monosulfide cross-links, which have higher thermal stability than di- and polysulfide cross-links. For optimum dispersion of small amounts of sulfur, some suppliers offer sulfur preblended with magnesium carbonate or in polymeric master batch forms.

Insoluble Sulfur. In natural rubber compounds, insoluble sulfur is used for adhesion to brass-coated wire, a necessary component in steel-belted radial tires. The adhesion of rubber to the brass-plated steel cord during vulcanization improves with high sulfur levels (~3.5%). Ordinary rhombic sulfur blooms at this dose level. Crystals of sulfur on the surface to be bonded destroy building tack and lead to premature failure of the tire. Rubber mixtures containing insoluble sulfur must be kept cool (<100°C) or the amorphous polymeric form converts to rhombic crystals.

Sulfur Donors. MBSS, DPTH, and the thiuram disulfides (see Table 2) are examples. The morpholine disulfide and caprolactam disulfide examples in Table 4 can also donate one atom of sulfur from their molecular structure for cross-linking purposes. Monosulfide cross-links provide better thermal stability than the sulfur–sulfur bonds in di- and polysulfide cross-links, which predominate when elemental sulfur is used.

Table 4. Properties of Cross-Linking Agents

Agent	CAS Registry Number	Mol wt	Density, Mg m ⁻³	Melting point, °C	Acute oral LD ₅₀ , mg kg ^a
Vulcanizing agents					
sulfur	[10544-50-0]	256	2.07	~95 ^b	
insoluble sulfur	[9035-99-8]	high ^c	1.95	~115	
4,4'-dithiodimorpholine	[103-34-4]	236	1.35	123–131	5,600
N,N'-caprolactam disulfide	[23847-08-7]	288	1.3	>120	3,620
Peroxides					
dicumyl peroxide	[80-43-3]	270	1.00		4,100
2,5-dimethyl-2,5-di(<i>t</i> -butylperoxy)hexane	[78-63-7]	290	0.87	liquid	>32,000
2,5-dimethyl-2,5-di(<i>t</i> -butylperoxy)hexyne	[1068-27-5]	286	0.89	liquid	1,850 (ip)
2,5-dimethyl-2,5-di-(benzoylperoxy)hexane	[2618-77-1]	386	1.23		
2,2'-bis(<i>t</i> -butylperoxy)-di- <i>iso</i> -propylbenzene	[25155-25-3]	338	0.93		>23,000
1,1-bis(<i>t</i> -butylperoxy)-3,3,5-trimethyl cyclo-hexane	[6731-36-8]	302	0.91	liquid	>13,000
<i>n</i> -butyl 4,4-bis(<i>t</i> -butyl-peroxy)valerate	[995-33-5]	334	0.95	liquid	5,000
<i>t</i> -butyl -perbenzoate	[614-45-9]	194	1.04	liquid	>3,600
benzoyl peroxide	[94-36-0]	242			>5,000
Hydroxy compounds					
<i>para</i> -quinone dioxime	[105-11-3]	138	1.40	215 dec	679
alkylphenol disulfide	[68555-98-6]		1.20	78–93	
methylolphenol–formaldehyde resin	[26678-93-3]	~1,000	1.05	60–70	
brominated alkylphenol formaldehyde resin	[112484-41-0]	~900	1.10	57–66	
Mercapto compounds					
2-mercapto-1,3,4-thia-diazole-5-thiobenzoate	[51988-14-8]	254	1.46		2,500
5,5'-thiobis-1,3,4-thia-diazole-2(3 <i>H</i>)thione	[7340-97-8]	266	1.90	172–177	930
trithiocyanuric acid	[638-16-4]	177	1.66	200 dec	>5,000 ^d
Diamino compounds					
hexamethylenediamine carbamate	[143-06-0]	160	1.28	152–155	2,875
N,N'-dicinnamylidene-1,6-hexanediamine	[140-73-8]	345	1.09	82–88	^d
4,4'-methylenebis-(cyclohexylamine)carbamate	[15484-34-1]	254	1.23	145–155	1,000
4,4'-methylenedianiline	[101-77-9]	200	1.15	70–80	200 ^e
Other difunctional cross-linking agents					
hexamethylene-1,6-bis-thiosulfate, disodium salt	[5719-73-3]	322	1.39	125 ^f	>5,000
1,3-bis(ditraconimido-methyl)benzene	[119462-56-5]	234	1.27	70–85	>2,000
N,N'- <i>meta</i> -phenylenebis-maleimide	[3006-93-7]	268	1.44	195–202	1,370

^a Toxicity data are for rats unless otherwise indicated.

^b Transition to monoclinic crystal form.

^c Approximately 200,000 molecular weight.

^d 4-h inhalation LC₅₀ > 5 mg /L.

^e Classified as IARC carcinogen.

^f Begins to lose water of hydration.

Peroxy Compounds. Peroxides give carbon–carbon cross-links, which provide rubber articles with the maximum resistance to heat, oxygen, and compression set. Resistance to compression set is important in applications such as gaskets and O-rings, which must maintain a seal by recovering their initial dimensions after deformation. Peroxides are beneficial in polymer blends or fully saturated polymers that cannot be cross-linked by other methods. Dialkyl peroxides are the principal type of peroxide used in the rubber industry. A prime example is dicumyl peroxide which sells for about \$9 /kg and is used in wire and cable insulation for good electrical properties in applications requiring water immersion. Peroxyketals find growing use for faster cross-linking and greater productivity. Benzoyl peroxide is used mainly in silicone.

Coagents are often used with peroxides to increase the state of cure. Some coagents, such as polybutadiene or multifunctional methacrylates, are used at high levels to form polymer grafts or interpenetrating networks. Other coagents such as triallyl cyanurate, triallyl trimellitate, and *meta*-phenylene bismaleimide are used at low levels to reduce the tendency of the polymer to degrade by chain scission.

The use of peroxides requires special handling precautions and compounding considerations. Compounds containing peroxides must be protected against exposure to air (oxygen) during the high temperature cross-linking process or chain scission and a tacky surface result. Some branched polymers such as butyl rubber, epichlorohydrin, and polypropylene undergo chain scission with peroxides and cannot be cross-linked by this method. Acidic materials heterolytically decompose peroxides without generating cross-links. Peroxy radical traps such as aromatic process oils (5) and many

antidegradants interfere with the cross-linking process, giving a lower state of cure (lower hardness and modulus). The standard antioxidant for peroxide cross-linking is polymerized 1,2-dihydro-2,2,4-trimethylquinoline.

Most peroxides of the appropriate thermal stability for rubber cross-linking are liquids or low melting solids which, in addition to the 100% active product, are offered adsorbed on inert, free-flowing carriers or in polymeric master batches for easy handling. Peroxides (qv) are oxidizing materials and should be stored away from other rubber chemicals and kept away from heat sources.

Multifunctional Hydroxy, Mercapto, and Amino Compounds. These are used to cross-link halogenated polymers. Depending on the lability of the halogen, the cross-linking agents can be capped to reduce reactivity or used in combination with accelerators to increase the rate of reaction. Benzoyl capping is common with hydroxy and mercapto compounds; forming the carbamate by reaction with one equivalent of carbon dioxide is used with diamines.

Quinone dioximes, alkylphenol disulfides, and phenol-formaldehyde reaction products are used to cross-link halobutyl rubbers. In some cases, nonhalogenated butyl rubber can be cross-linked by these materials if there is some other source of halogen in the formulation. Alkylphenol disulfides are used in halobutyl innerliners for tires. Methylol phenol-formaldehyde resins are used for heat resistance in tire curing bladders. Bisphenols, accelerated by phosphonium salts, are used to cross-link fluorocarbon rubbers.

2,5-Dimercapto-1,3,4-thiadiazole derivatives, accelerated by amines, are used to cross-link chlorinated polyethylene. Polyisobutylene containing brominated *para*-methylstyrene cure functionality can be cross-linked in polymer blends with dimercapto-1,3,4-thiadiazole derivatives accelerated with thiuram disulfides. Trithiocyanuric acid is suggested for use in polyacrylates containing a chlorine cure site and in epichlorohydrin rubbers.

Diamine curatives were the first cross-linking agents for fluorocarbon rubbers. They are corrosive to mild steel molds and have been replaced in many applications by the bisphenol or other more recent cure systems. Nevertheless, some diamines are still used for food-contact applications of fluorocarbon rubbers and in zinc-free cures of halobutyl rubbers for pharmaceutical stoppers. Methylene dianiline and triethylene tetramine are cross-linking agents for ethylene-acrylic elastomers.

Other Difunctional Cross-Linking Agents. These are used in certain polymers and to provide specialized properties. Both hexamethylene-1,6-bisthiosulfate and 1,3-bis(ditraconimidomethyl)benzene are reported to provide long, flexible cross-links for good resistance to fatigue by flexing without the poorer thermal stability inherent with polysulfide cross-links. *meta*-Phenylenebismaleimide is a coagent used to increase the state of cure with peroxides. The maleimide can also contribute to cross-linking in chlorosulfonated polyethylene and provide thermal stability in sulfur cures of highly unsaturated rubbers.

Activators and Retarders

Table 5 shows common properties of activators and retarders.

Table 5. Properties of Activators and Retarders

Activator-retarder	CAS Registry Number	Mol wt	Density, Mg m ⁻³	Melting point, °C	Acute oral LD ₅₀ , mg kg ⁻¹ ^a
Activators					
zinc oxide	[1314-13-2]	81	5.61	sublimes	5,400
stearic acid (octa-decanoic acid)	[57-11-4]	284	0.85	55–70	21.5 (iv)
zinc stearate	[557-05-1]	632	1.05	124–126	
lauric acid (dodeca-noic acid)	[143-07-7]	200	0.87	40–50	131 (iv) ^b
zinc laurate	[68242-42-0]	364	1.18	107–114	
zinc 2-ethylhexoate	[136-53-8]	352	1.11	liquid	3,550
Metal oxides					
magnesium oxide	[1309-48-4]	40	3.5	2,800	
litharge (lead monoxide, PbO)	[1317-36-8]	223	9.53	888	430 (ip) ^c
red lead (lead tetroxide, Pb ₃ O ₄)	[1314-41-6]	686	9.1	500 dec	200 (ip) ^d
synthetic hydrotalcite ^e	[12304-65-3]		2.09		
calcium oxide	[1305-78-8]	56	3.33	2,572	
zinc carbonate (hydrozincite)	[12122-17-7]	125	3.5		
Acidic retarders					
phthalic anhydride	[85-44-9]	148	1.48	125–130	4,020
benzoic acid	[65-85-0]	122	1.32	122 ^f	2,530
salicylic acid	[69-72-7]	138	1.44	157–159 ^g	890
High performance retarders					
N-(cyclohexyl-thio)phthalimide	[17796-82-6]	260	1.3	90–95	2,600
sulfonamide derivative	[2280-49-1]		1.68	110	>2,500
tetra- <i>iso</i> -butyl thiuram monosulfide		377	1.08	60–68	

^a Toxicity data are for rats unless otherwise indicated.
^b For mice.
^c For rats (poisonous).
^d For guinea pigs (poisonous).
^e Mg_{1-x}Al_x(OH)₂CO₂H₂O.
^f Begins to sublime at ~100°C.
^g Sublimes at 76°C.

Zinc oxide is a common activator in rubber formulations. It reacts during vulcanization with most accelerators to form the highly active zinc salt. A preceding reaction with stearic acid forms the hydrocarbon-soluble zinc stearate and liberates water before the onset of cross-linking (6). In cures at atmospheric pressure, such as continuous extrusions, the prereacted zinc stearate can be used to avoid the evolution of water that would otherwise lead to undesirable porosity. In these applications, calcium oxide is also added as a desiccant to remove water from all sources.

For increased solubility to prevent bloom, shorter-chain carboxylic acids or zinc carboxylates can be substituted. The use of chain-branched carboxylic acids reduces the tendency for the formulations to lose sulfur cross-links or revert upon prolonged heating (7). Translucent articles such as crepe soles can use a zinc carboxylate or employ zinc carbonate as a transparent zinc oxide.

Magnesium oxide is a typical acid scavenger for chlorinated rubbers. Compounds containing zinc oxide or magnesium oxide may tend to swell upon immersion in water. These inorganic salts have some water solubility and osmotic pressure causes the vulcanizates to imbibe water to equalize pressure (8,9). As such, vulcanizates tend to swell more in fresh (distilled) water than in salt water. To minimize water swell, insoluble salts such as lead oxides can be substituted. Because of the health concerns associated with lead, there is much rubber industry interest in other acid acceptors, such as synthetic

hydrotalcite.

Retarders were originally arenecarboxylic acids. These acidic materials not only delay the onset of cross-linking but also slow the cross-linking reaction itself. The acidic retarders do not function well in black-filled compounds because of the high pH of furnace blacks. Another type of retarder, *N*-nitroso diphenylamine [86-30-6], was used for many years in black-filled compounds. This product disappeared when it was recognized that it trans-nitrosated volatile amines to give a several-fold increase in airborne nitrosamines. U.S. production peaked in 1974 at about 1.6 million kg.

Modern retarders are designed to scavenge the most reactive accelerator precursors and are known as prevulcanization inhibitors. This technique delays the onset of cure without affecting the speed of the main cross-linking reaction. *N*-(Cyclohexylthio)phthalimide was the first high performance retarder to find wide acceptance (10). This retarder sells for about \$11 kg and its use has grown with the switch from sulfenamides based on secondary amines to those based on primary amines which give less delayed action. The sulfonamide provides similar performance whereas the developmental thiuram monosulfide provides a modest delay in cure onset along with a reduction in cure time. All these high performance retarders function principally with sulfenamide accelerators.

Overview of Vulcanization Chemistry

Sulfur vulcanization leads to a variety of cross-link structures as shown in Figure 1. All the sulfur does not result in cross-links; some of it remains as pendent accelerator polysulfide groups and internal cyclic polysulfides. These alternative structures do not contribute to load bearing or strength properties and are more prevalent in unaccelerated or weakly accelerated vulcanization systems. Additional heating can also reduce the polysulfide rank of the cross-links. In some elastomers, this leads to a larger number of cross-links. However, in natural rubber or its synthetic polyisoprene equivalent, the overall result is a loss of cross-links, especially at temperatures over 160°C.

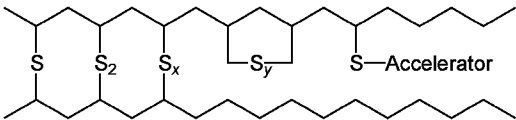


Fig. 1. Structures formed during sulfur vulcanization of elastomers.

Although it has been studied since the 1950s, the exact mechanism of accelerated sulfur vulcanization remains unresolved, including whether it proceeds by a radical or ionic process. One review (11) suggests that the nature of the reaction may change, depending on the polarity of the particular polymer (solvent) and whether or not zinc oxide is present. Sulfur vulcanization employs a combination of zinc oxide, fatty acid, sulfur, and at least one accelerator. These materials react as shown in Figure 2 to form a complex (I) in which the eight-member sulfur ring has been opened (12). The complex then abstracts a hydrogen on an alpha-carbon position to the double bond in the polymer to form a rubber-bound pendent accelerator polysulfide (II), a cross-link precursor. In Step 3, a zinc complex and an adjacent polymer chain react with the pendent polysulfide to form the polysulfide cross-link while regenerating the zinc–accelerator complex. The cycle is continued until all the sulfur is consumed.

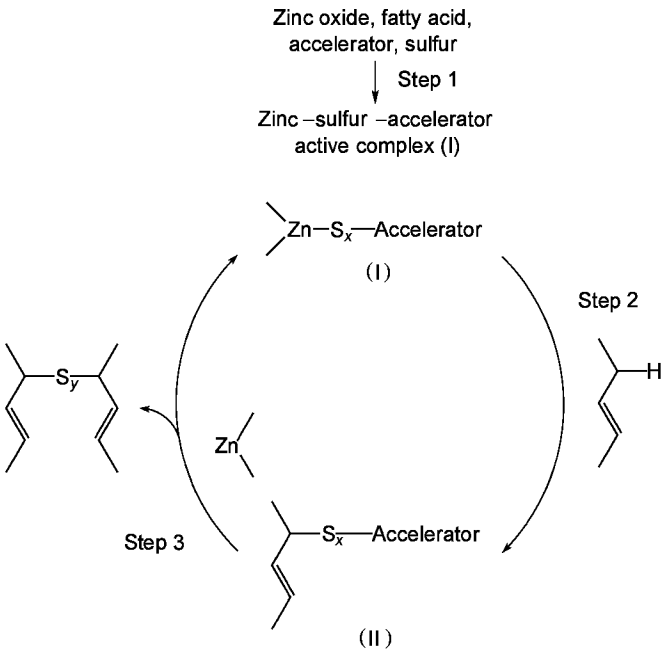
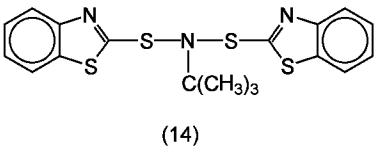


Fig. 2. Cyclic nature of vulcanization process (12).

In the absence of zinc oxide, cross-linking proceeds through an accelerator polysulfide. With TBSI (14) and other sulfenamides, the accelerator decomposes upon heating during the induction period (before cross-linking) as shown in Figure 2 (13).



The decomposition generates amine and MBTS. Small quantities of MBT and benzothiazyl (Bt) polysulfides form just before the onset of cross-linking. Sulfur disappears and large quantities of MBT are formed during the main cross-linking reaction. The development of cross-links is shown as the rheometer line (curve C) in Figure 3.

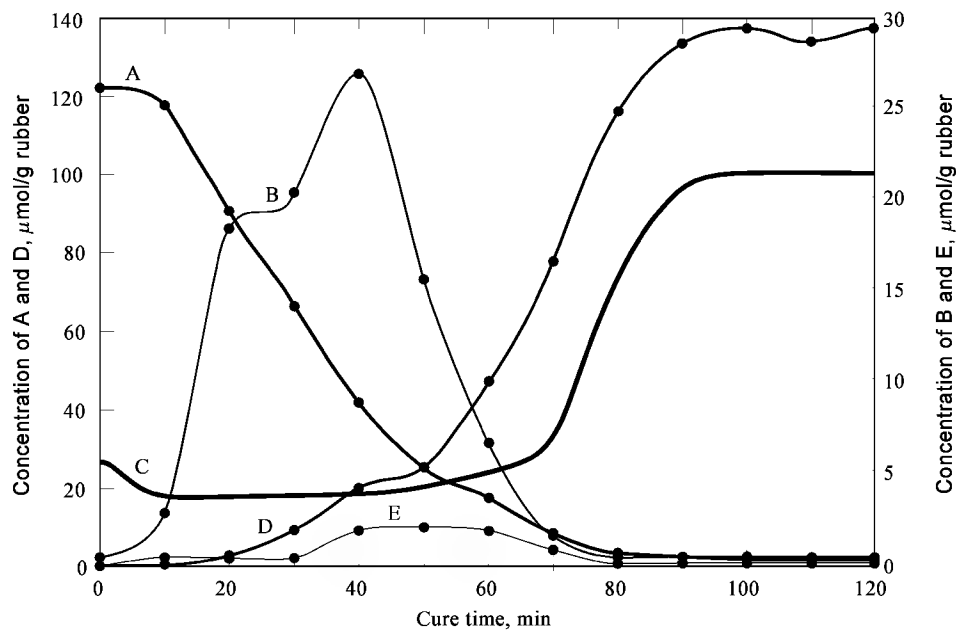
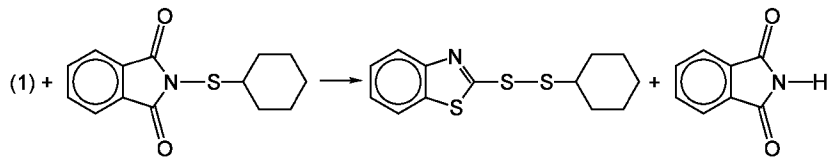


Fig. 3. Fate of accelerator components during cure, where A is TBSI; B, MBTS; C, development of cross-links (rheometer); D, MBT; and E, BtS₃Bt (13).

The inhibition chemistry has been extensively studied. As shown in equation 1, the prevulcanization inhibitor (retarder) reacts with MBT (1) before it can form polysulfides.



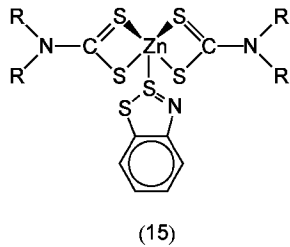
For optimum cross-linking efficiency, a combination of accelerators is used. Table 6 shows the increase in load-bearing capability of vulcanizates based on different rubbers as the ratio of two accelerators is changed (14). The term 300% modulus represents the strain at 300% stress and is not a true modulus because rubber gives nonlinear stress–strain behavior. For polymers with primary allylic carbon atoms, the use of two accelerators gives significantly higher 300% modulus than either accelerator used alone. When the rubber polymer consists of secondary allylic carbon atoms, the modulus is level until the sulfenamide OBTS becomes the principle accelerator.

Table 6. Effect of Accelerator Ratio on 300% Modulus, MPa^a

OTOS, % OBTS, %	100 0	67 33	50 50	0 100	Total accelerator ^b
Rubbers based on CH ₃ –C=C–					
natural rubber (polyisoprene)	15.4	19.7	20.0	15.7	1.44
butyl rubber (poly- <i>iso</i> -butylene)	3.6	3.9	3.8	3.2	3.20
EPDM					
1,4-hexadiene cure site	2.9	3.6	3.9	3.1	3.0
ethylenenorbornene cure site	4.8	4.9	4.8	4.2	2.0
Rubbers based on –CH ₂ –C=C–					
EPDM, dicyclopentadiene cure site	3.0	3.1	3.0	2.2	2.0
polybutadiene	12.4	11.8	12.2	8.4	1.0
butadiene–styrene copolymer ^c	8.2	7.9	7.3	6.2	0.9
butadiene–acrylonitrile copolymer	12.0	11.5	11.2	8.4	1.25

^a Ref. 14. To convert MPa to psi, multiply by 145.
^b Parts per 100 parts of rubber polymer (phr).
^c Blended with polybutadiene.

The optimum modulus occurs at about a 2:1 weight ratio of OTOS to OBTS. Similar optimums have been observed with other accelerator combinations. The examples shown in Figure 4 are calculated from regression equations developed from designed experiments in a black-filled natural rubber compound. On a molar basis, the synergistic accelerator complex appears to consist of two dithiocarbamate ligands and one mercaptobenzothiazole moiety, as shown in structure (15) (14).



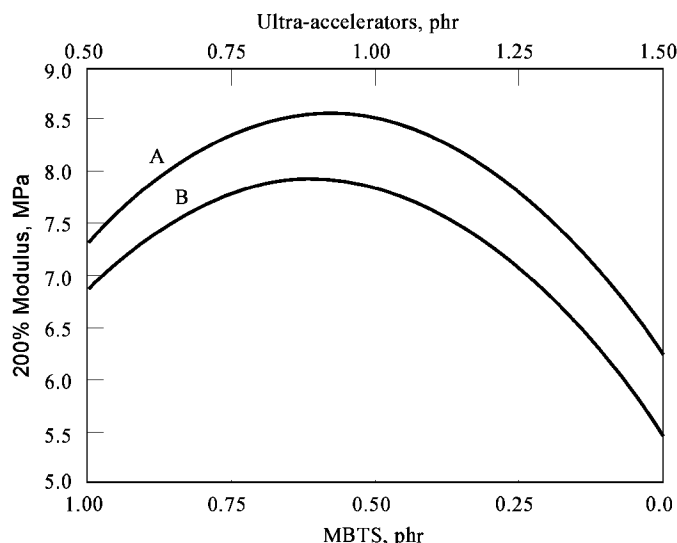


Fig. 4. Accelerator synergism where A is tetra-*iso*-butyl thiuram disulfide and B is zinc di-*iso*-butyl dithiocarbamate. To convert MPa to psi, multiply by 145.

Processing Agents

Natural rubber must be reduced in viscosity in order to obtain workable compounds. Many different chemical peptizers have been employed over the years for this purpose, including arenemercaptans, arenesulfonic acids, pentachlorothiophenol [133-49-3] or its zinc salt, and dithiobisbenzanilide [135-57-9] or its zinc salt. Dithiobisbenzanilide with an activator and clay diluent is the preferred peptizing agent for natural and synthetic rubbers. The viscosity of natural rubber and synthetic polyisoprene can be reduced by mechanical shear alone, but using a peptizer makes the viscosity reduction during mixing less sensitive to variations in time and temperature, providing uniformity in viscosity from batch to batch.

Processing aids also assist flow but do so by physical rather than chemical means. Processing aids may be separated into two general categories: external and internal. External processing aids have a polar functionality at one end of the molecule and a long-chain, oleophilic tail. These materials tend to migrate to surfaces where the polar head becomes adsorbed and the oleophilic tail allows the hydrocarbon polymer to slide along the coated surface with less effort. The surface can be either the metal walls of the processing equipment or the filler particles introduced into the rubber matrix. Internal processing aids such as oils or waxes increase melt flow. They reduce the viscosity of the rubber matrix by allowing polymer chains to slide past one another with greater ease. Many high performance processing aids are proprietary mixtures of materials with both internal and external functions.

Homogenizing agents are higher molecular weight versions of external processing aids. They are useful in polymer blends to reduce the domain size of the dispersed polymer phase. Homogenizing agents can be particularly useful in blending polymers having dissimilar solubility parameters.

Vulcanized vegetable oils are processing aids that flow under shear but not under heat. They are made by heating unsaturated soybean or rapeseed oil with either sulfur monochloride, S_2Cl_2 , or elemental sulfur. The products cross-linked with S_2Cl_2 are white powders. They are used as a main component of erasers to allow greater ability to lift a pencil mark without smearing (noodling). The sulfur-vulcanized products are dark brown solids, used in soft roll compounds to help hold in liquid plasticizers and assist the grinding of the rubber-covered roller to concentricity.

Other processing aids are tackifiers or antitack agents. Tackifiers are resinous products which help a multilayered article stick together before it is vulcanized. Tackifiers may be aliphatic or aromatic materials derived from coal, petroleum, or renewable sources. A variety of additives are used to keep unvulcanized sheets of rubber from sticking together and can be applied by dusting, dipping, or spraying.

Other Rubber Chemicals

Blowing agents are used to create cellular products and have been reviewed (15). Closed-cell sponge is used for minimal water absorption in applications such as wet suits or automotive weather-strip. For closed-cell sponge, three nitrogen-releasing blowing agents are used: azodicarbonamide [123-77-3], *p,p'*-oxybis(benzenesulfonyl hydrazide) [80-51-3], and dinitrosopentamethylenetetramine [101-25-7]. Open-cell sponge is used for cushioning when the product must deform without much effort, eg, rug carpet underlay. The basic blowing agent for open-cell sponge is sodium bicarbonate [144-55-8]. Many other rubber chemicals in the mixed compound can affect the temperature and rate at which blowing agents decompose to release gas. Careful compound development and process control are needed to match the gas generation with the cross-linking reaction.

Adhesion promoters are used to bond rubber to brass-plated steel cord in radial tires or conveyor belting. A gradual decrease in stiffness from steel to rubber is desired. A thin layer of a harder rubber compound is applied to the steel to aid bonding. In addition to higher sulfur levels for increased cross-link density, the bonding layer contains resin-forming additives. One resin-forming component is a methylene acceptor such as resorcinol [108-46-3] or a thermoplastic resorcinol-formaldehyde condensate. The other component is formaldehyde or a methylene donor such as hexamethylenetetramine or hexamethoxymethylmelamine [3089-11-0]. Optionally, cobalt naphthenate or other salts can be included to modify the rate at which sulfur reacts with the brass coating (16).

Rubber processed in latex form accounts for about 10% of new rubber consumption. Rubber latex is a liquid, oil-in-water emulsion which is used to make foam or thin-walled rubber articles. The same accelerators and antidegradants used in dry rubber are used in latex, with longer-chain versions preferred for greater oil solubility. To prepare these and other additives for addition to latex, they must be predispersed in water and the surface of the powder or oil droplet coated with a surface-active agent to prevent destabilization (coagulation) of the latex.

Dispersing agents such as sodium salts of polymerized alkylarenesulfonates [9084-06-4] or [8061-51-6] are surface-active agents which only weakly reduce the surface tension. Wetting agents such as sodium lauryl sulfate [151-21-3] more strongly reduce surface tension. Dispersing agents are used when a large quantity of surface area must be treated with surfactant without the mixture entrapping bubbles; wetting agents are used to adjust the amount of fabric penetration or the flow of the latex mixture. Thickeners are used to prevent the settling of additives and can be organic polyelectrolytes or inorganic smectites [12199-37-0]. Colloidal stabilizers such as casein are also used but are falling into disfavor because of the desire to avoid protein which can cause anaphylactic shock in sensitive individuals. Various antifoams prevent air entrapment and preservatives help avoid growth of microbes that could create slime or odors (see INDUSTRIAL ANTIMICROBIAL AGENTS). Once formulated with additives, the rubber article is formed by adding acid-generating gellants to natural rubber or calcium nitrate [10124-37-5] to polychloroprene.

Flame retardants such as α -alumina trihydrate [14762-49-3] can be added to latex-based foamed carpet backing; a combination of antimony oxide [1309-64-4] and chlorinated paraffins is used in dry rubber.

Organofunctional silanes are used to promote polymer-to-filler bonding with clay or silica fillers. Vinyl silanes are used in peroxide-cured wire insulation to promote stronger bonding with calcined clay fillers. Mercapto silanes are used to treat kaolin clay in sulfur-cured compounds.

Consumer articles often use colorants (qv), reodorants, or finishing agents. Carbon black (qv) provides the best technological properties for industrial applications, so most rubber articles are black. Red iron oxide or other inorganic pigments are used to color mineral-filled articles. Organic

coloring materials provide brighter colors, although special compounding or processing may be required to maintain the color. For example, sulfur cure systems quench fluorescent pigments so peroxide cures are needed. Reodorants can be used to provide a more pleasing aroma for consumer articles. Finishing agents can be applied for a more attractive surface appearance.

BIBLIOGRAPHY

"Rubber Chemicals" in *ECT* 1st ed., Vol. 11, pp. 870–892, by R. A. Mathes and A. E. Juve, The B. F. Goodrich Research Center; in *ECT* 2nd ed., Vol. 17, pp. 509–542, by F. Shaver, The B. F. Goodrich Co.; in *ECT* 3rd ed., Vol. 20, pp. 337–364, by R. Taylor and P. N. Son, The B. F. Goodrich Co.

1. *Synthetic Organic Chemicals*, United States International Trade Commission Reports, Government Printing Office, Washington, D.C., 1965–1994.
2. **U.S. Pat. 3,633** (June 15, 1844), C. Goodyear.
3. G. Oenslager and Cf. W. C. Greer, *J. Ind. Eng. Chem.* **14**, 373 (1922).
4. M. A. Thomson, C. R. Green, R. B. Balodis, and co-workers, "N-Nitrosamines", in *Patty's Industrial Hygiene and Toxicology*, 4th ed., Vol. 2, John Wiley & Sons, Inc., New York, 1993, Part A, Chapt. 11.
5. K. L. Chasey, *Rubber Chem. Tech.* **65**, 385 (1992).
6. A. B. Sullivan, C. J. Hann, and G. H. Kuhls, "Vulcanization Chemistry—Fate of Elemental Sulfur and Accelerator during Scorch Delay as Studied by Modern HPLC", Paper No. 9, presented at the *ACS Rubber Division Meeting, Toronto, Canada, May 21–24, 1991*, American Chemical Society, Washington, D.C., 1991.
7. J. Vander Kooi, *Rubber Plast. News*, 17–19 (May 23, 1994).
8. D. C. Edwards, *Rubber Chem. Technol.* **39**, 581 (1966).
9. A. Hallenbeck and S. W. Schmitt, *Compounding Neoprene for Water Resistance*, NP-520.1 (E-27911), E.I. du Pont de Nemours and Co., Inc., Wilmington, Del., 1987.
10. **U.S. Pats. 3,427,319** (Feb. 11, 1969), **3,546,185** (Dec. 8, 1970), **3,752,824** (Aug. 14, 1973), and **3,855,262** (Dec. 17, 1974), A. Y. Coran and E. Kerwood (to Monsanto Co.).
11. M. R. Kvejsa and J. L. Koenig, *Rubber Chem. Technol.* **66**, 376 (1993).
12. R. W. Layer, *Rubber Chem. Technol.* **65**, 211 (1992).
13. C. J. Hann, A. B. Sullivan, B. C. Host, and G. H. Kuhls, Jr., *Rubber Chem. Technol.* **67**, 76 (1994).
14. R. W. Layer, *Rubber Chem. Technol.* **62**, 124 (1989).
15. D. G. Rowland, *Rubber Chem. Technol.* **66**, 463 (1993).
16. W. J. van Ooij, *Rubber Chem. Technol.* **57**, 421 (1984).

General References

M. Morton, ed., *Rubber Technology*, 3rd ed., Van Nostrand Reinhold, New York, 1987. An introductory textbook for novices.

H. Long, ed., *Basic Compounding and Processing of Rubber*, ACS Rubber Division, Washington, D.C., 1985. An intermediate-level textbook focusing on basic manufacturing processes for rubber articles.

F. R. Eirich, ed., *Science and Technology of Rubber*, 2nd ed., Academic Press, Inc., New York, 1994. An advanced treatise.

R. F. Ohm, ed., *The Vanderbilt Rubber Handbook*, 13th ed., R. T. Vanderbilt Co., Inc., Norwalk, Conn., 1990. Contains general information for novices and practical reference formulations for experienced compounders.

R. F. Mausser, ed., *The Vanderbilt Latex Handbook*, 3rd ed., R. T. Vanderbilt Co., Inc., Norwalk, Conn., 1987. Contains introductory and reference information for people engaged in the latex branch of the rubber industry.

D. R. Smith, ed., *Rubber World Magazine's 1995 Blue Book*, Lippincott & Peto, Akron, Ohio, 1995. An annual compilation of materials, compounding ingredients, machinery, and services for the rubber industry. Available on CD-ROM in combination with R. F. Ohm, ed., *The Vanderbilt Rubber Handbook*, 14th ed., 1995.

Robert F. Ohm
R. T. Vanderbilt Company, Inc.

RUBBER COMPOUNDING

Rubber compounding combines the art and science of selecting the best combination of elastomer and other ingredients that meet the processing, cost, and performance quality, and cost requirements for rubber goods made and used in commerce. There is a wide variety of elastomers and ingredients that are available in making rubber goods, which include all of the following types of products: tires, inner tubes, retreaded tires, footwear, rubber rolls, hose, belts, weatherstripping, O-rings, seals, diaphragms, tubing, rubber and latex gloves, ball bladders, condoms, bumpers, and numerous other products. The purpose of this article is to review the logic and rationale used in designing many common types of rubber products. The creation of the combination of components that produce a usable rubber compound involve many classical disciplines, including chemistry, physics, rheology, and mechanics of polymers and understanding how compounding components interact. In many cases classical science can predict the interaction effects and project the properties expected from rubber compounds. However, discovery and development of new materials that interact with the rubber compound to produce useful and beneficial properties are appearing regularly. Also environmental concerns are creating an interest in utilizing used rubber products in virgin rubber compounds (see RECYCLING, RUBBER).

Compounding Hierarchy. In designing or making the selection of ingredients for use in a rubber compound the compounder relies on experience, education, training, and various information sources such as supplier data and technical reports. Ingredients are generally selected in the following order: (1) polymer (natural or synthetic rubber), (2) vulcanization system (curing agent, accelerator(s), or coagent), (3) reinforcing agent and fillers, (4) plasticizers or oil, (5) antioxidant and antiozonant, (6) bonding agent or adhesive, and (7) tackifier if needed.

Economics and price of the final article often dictate a specific type of rubber that can be used. The expected usable life for the product is controlled by many factors including end customer awareness, competitive situation in the marketplace, safety, reliability, and other factors. Rubber is almost always used as a functional part of another system. For example tires, hoses, belts, O-rings, and numerous rubber components are used in manufacturing automobiles and trucks. The overall life of the vehicle as well as its performance level often control or directly relate to the service life or quality level of the rubber parts.

Elastomers Used in Rubber Compounding

The principal component of a rubber compound is the elastomer or blend of elastomers chosen for a specific component application. There are 25–30 different chemical classifications of elastomers; six of these classes represent over 90% of all elastomers used (see ELASTOMERS, SYNTHETIC).

According to the International Institute of Synthetic Producers (IISRP) there were 3.8×10^6 t of elastomer consumed in the United States in 1994, 27% natural rubber, and 73% synthetic rubber. Elastomer growth in the United States is ~1.5% and closely follows GDP growth (1).

Natural rubber comes generally from southeast Asia. Synthetic rubbers are produced from monomers obtained from the cracking and refining of petroleum (qv). The most common monomers are styrene, butadiene, isobutylene, isoprene, ethylene, propylene, and acrylonitrile. There are numerous others for specialty elastomers which include acrylics, chlorosulfonated polyethylene, chlorinated polyethylene, epichlorohydrin, ethylene–acrylic, ethylene octene rubber, ethylene–propylene rubber, fluoroelastomers, polynorbornene, polysulfides, silicone, thermoplastic elastomers, urethanes, and ethylene–vinyl acetate.

Elastomers are classified by their heat resistance and resistance to swelling by IRM 903. The SAE J200 system is employed in doing the fundamental classification. Properties for evaluating elastomers include hardness, tensile strength, change in tensile strength, elongation and hardness, change in volume, tensile strength, and hardness after exposure to IRM 903 oil. Figure 1 shows an overview of the primary classes of elastomers based on this system. Table 1 summarizes other properties of elastomers.

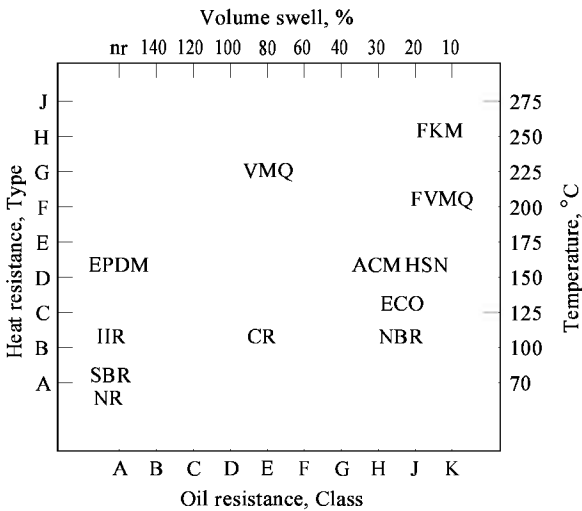


Fig. 1. SAE J200 Classification system for ASTM No. 3 oil where in volume swell nr = no requirement. EPDM is ethylene–propylene–diene monomer; IIR, butyl rubber; SBR, styrene–butadiene rubber; NR, natural rubber; VMQ, methyl vinyl silicone; CR, chloroprene; FKM, fluoroelastomer; FVMQ, fluorovinyl methyl silicone; ACM, acrylic elastomers; HSN, hydrogenated nitrile; ECO, epichlorohydrin; and NBR, nitrile rubber.

Table 1. Physical Properties of Elastomers^a

Elastomer ^b	Specific gravity	Hardness, Shore A	Tensile strength, ^c MPa ^d	Elongation, ^c %	Resilience ^e	Compression set ^e	Impermeability to gases ^e
NR	0.93	30–100	27.6	750	E	G	F
isoprene	0.92	30–100	24.1	750	E	F	F
SBR	0.94	35–100	20.7	600	G	G	F

butyl	0.92	30–90	17.2	700	P–F	P–F	E
butadiene	0.91	45–80	17.2	500	E	F	F
EPDM	0.86	30–90	20.7	600	G	G	F
chloroprene	1.23	35–95	20.7	600	G–E	F–G	F–G
nitrile	1.00	30–100	20.7	600	F–G	G	G
thiokol	1.25–1.35	20–80	10.3	450	P–F	P–F	E
urethane	1.02–1.25	55–100	55.2	750	F–E	G–E	P–F
silicone	0.98–1.60	25–90	10.3	800	F–G	G–E	P–F
CSM	1.12–1.28	40–95	20.7	600	F–G	F	G
acrylic	1.09	40–90	13.8	400	F–G	G	F–G
fluorocarbon	1.85	55–95	20.7	450	F	G–E	G–E
epichlorohydrin	1.27–1.36	40–90	17.2	400	F–G	F	E
chlorinated PE	1.16–1.25	45–95	24.1	600	F–G	F–G	G
cross-linked PE	0.92	90 +	20.7	500	P	F	G

^a Ref. 2.

^b NR = natural rubber; SBR = styrene–butadiene rubber; EPDM = ethylene–propylene–diene monomer; CSM = chlorosulfonated polyethylene; and PE = polyethylene.

^c Maximum value at room temperature.

^d To convert MPa to psi, multiply by 145.

^e E = excellent; G = good; F = fair; and P = poor.

Natural Rubber. To obtain natural rubber (NR), the *Hevea brasiliensis* tree is tapped for its sap. The off-white sap is collected and coagulated. This process produces a high molecular weight substance which is natural rubber. The principal producing countries are Malaysia, Indonesia, Thailand, India, China, and Sri Lanka (see RUBBER, NATURAL).

There are several systems that define the quality and uniformity of natural rubber. One system of grading natural rubber is based on form and visual observation of color and cleanliness. This is known as the International Natural Rubber Specification. The principal types and grades are as follows. There are five other types of rubber classified by this system and many other grades not listed here.

Type	Grade	Example
1	ribbed smoked sheet	RSS 1 or 3
2	pale and white crepe	1, 2, 3
3	estate brown crepe	1X, 2X

Malaysian and Indonesian natural rubber growers have established a system based on technical characteristics. The standard Malaysian specification scheme can be found in Reference 3.

In addition to the solid form of natural rubber it is available as a solid suspended in water, known as latex. Synthetic rubbers are also available in latex form. Latex has become an important commodity used in the manufacture of dipped goods for health and disease protection. The principal uses of natural rubber are as follows: tires and retreading, 70° o; latex (gloves, balloons), 12° o; mechanical goods, 9° o; load-bearing components, 4° o; and other, 5° o.

Styrene–Butadiene Rubber. Styrene–butadiene rubber (SBR) is made either as an emulsion or solution product. Emulsion SBR is the older of the two forms and less expensive. However, it is more difficult to control polymer microstructure in this process and the emulsion form is not as pure as the solution form. Solution forms are finding widespread use in original equipment passenger tires (see STYRENE BUTADIENE RUBBER). The principal use of SBR is in the tire industry as a major component of all passenger tires and a significant portion of all tire products, except aircraft and large off-the-road tires. SBR uses in North America are tires, 75° o; automotive mechanical goods, 8° o; and other, 17° o (4). Important properties of SBR are relatively low cost, good physical properties with the proper reinforcing agent, good abrasion, and skid resistance. A deficiency of SBR is its tendency to have fairly high heat generation.

Polybutadiene. The many forms that can result from the polymerization of butadiene, depending on the catalysts used, include high cis, medium cis, low cis, and high vinyl polybutadiene (PBD) (see ELASTOMERS, SYNTHETIC POLYBUTADIENE).

The property of polybutadiene of most interest to the rubber compounder is excellent abrasion resistance coupled with excellent resilience. The polymer has very high rebound and low heat generation. With a few exceptions, such as the core of solid golf balls, the polymer is blended with other polymers to take advantage of its excellent abrasion and rebound. Uses in North America are as follows: tires, 500,000 t (74° o); plastic modification, 136,000 t (20° o); golf balls, 13,600 t (2° o); and miscellaneous nontire, 27,000 t (4° o) (4). The principal grades of polybutadiene used in the rubber and plastic industry are shown in Table 2.

Table 2. Commercially Available Elastomeric Polybutadienes

Producer	Location	Trade name	Catalyst type ^a
<i>High cis (>90% cis-1,4)</i>			
American Synthetic Rubber Corp.	North America	Cisdene	Z-N
Australian Synthetic Rubber Co.	Australia	Austrapol	Co–Z-N
Bunawerke HbIs GmbH	Europe	Buna	Co–Z-N, Ti–Z-N, Nd–rare earth
Enichem	Europe	Europene cis, Neocis	Ti–Z-N, Nd–rare earth
Goodyear	North America	Budene	Co–Z-N
Indian Petrochemical Corp.	India	Cisamer	Co–Z-N
JSR	Japan	JSR	Ni–Z-N
Kumho Petrochemical	Korea	Kosyn	Ni–Z-N
Nippon Zeon	Japan	Nipol	Co–Z-N
Petkim	Europe	Petkacuk	Co–Z-N
Bayer	North America	Taktene	Co–Z-N
Shell Chemical SA	Europe	Cariflex	Co–Z-N
Taiwan Synthetic Rubber	Taiwan	Taipol	Co–Z-N
UBE Industries	Japan	Ubepol	Co–Z-N
<i>Low cis (<40% cis-1,4)</i>			

Asahi Chemical Co., Ltd.	Japan	Asadene	Li-alkyl
Bunawerke H&S GmbH	Europe	Buna	Li-alkyl
Calatrava	Europe	Calprene	Li-alkyl
Coperbo	South America	Coperflex	Li-alkyl
Enichem	Europe	Intene	Li-alkyl
Finaprene NV	Europe	Finaprene	Li-alkyl
Firestone Synthetic Rubber	North America	Diene	Li-alkyl
VED Chemischewerke buna	Europe	Buna	Li-alkyl
Negromex	Mexico	Solprene	Li-alkyl
Polysar	North America	Taktene	Li-alkyl
	<i>Vinyl butadiene rubber</i> ^b		
Bunawerke H&S GmbH	Europe	Buna	
Enichem	Europe	Intolene	
Goodyear	North America	Budene	
Nippon Zeon	Japan	Nipol	
Bayer	North America	Polysar VBR	

^a Z-N Ziegler-Natta.

^b Vinyl level for Buna and Intolene, medium–high; Budene, medium; and Nipol and Polysar VBR, high.

Ethylene–Propylene Rubber. Ethylene and propylene copolymerize to produce a wide range of elastomeric and thermoplastic products. Often a third monomer such dicyclopentadiene, hexadiene, or ethylene norbornene is incorporated at 2–12% into the polymer backbone and leads to the designation ethylene–propylene–diene monomer (EPDM) rubber (see ELASTOMERS, SYNTHETIC ETHYLENE PROPYLENE DIENE RUBBER). The third monomer introduces sites of unsaturation that allow vulcanization by conventional sulfur cures. At high levels of third monomer it is possible to achieve cure rates that are equivalent to conventional rubbers such as SBR and PBD. Ethylene–propylene rubber (EPR) requires peroxide vulcanization.

Variation of the ethylene–propylene ratio changes the viscoelastic properties of the rubber. The hardness can be varied widely by the filler and plasticizer level. In addition EPDM can be highly extended with black, nonblack fillers, and oils. At high levels of fillers, EPDM can display very good resistance to swelling by the new IRM 903 oil. The principal advantageous properties of EPR and EPDM are excellent weathering, good heat and water resistance, good dielectric strength, and low temperature brittleness. As a result EPDM is used in a wide variety of rubber goods. A rough breakdown of the 200 + million kilograms of EPR and EPDM consumed in the North American market is as follows: roofing, 20% ; plastic modification, TPOs, and TPVs, 17% ; weatherstripping, 15% ; hoses and belts, 14% ; oil viscosity modification, 10% ; wire and cable, 6% ; molded goods, 5% ; tires, 3% ; carpet backing, 2% ; and other, 9% (5).

EPR and EPDM have been made by either solution or emulsion polymerization processes. More recently a new process involving gas-phase polymerization and metallocene catalysts promises to capture large shares of these markets. These new polymers will be especially attractive in automotive applications and wire and cable where their favorable pricing should be welcome.

There are five North American manufacturers of EPDM: Uniroyal Chemical, Exxon, DuPont, DSM, and Bayer. In addition, Union Carbide and Dow are manufacturing the new gas-phase polymers. A new joint agreement between Dow and DuPont will offer these two companies some competitive advantages in this market. Trade names of commercial ethylene–propylene polymers include Bayer's Epsyn, Exxon's Vistalon, DuPont's Nordel, and Uniroyal's Royalene.

Butyl and Halobutyl Rubber. Butyl rubber is made by the polymerization of isobutylene; a small amount of isoprene is added to provide sites for curing. It is designated IIR because of these monomers. Halogenation of butyl rubber with bromine or chlorine increases the reaction rate for vulcanization and laminates or blends of halobutyl are feasible for production of rubber goods. It is estimated that of the ~100 million kg of butyl (IIR) and halobutyl (HIIR) rubber in North America, over 90% is used in tire applications. The halogenated polymer is used in the innerliner of tubeless tires. Butyl rubber is used to make innertubes and curing bladders. The two major suppliers of butyl and halobutyl polymers in North America are Exxon and Bayer (see ELASTOMERS, SYNTHETIC BUTYL RUBBER).

Butyl polymers are about 8–10 times more resistant to air permeability compared to natural rubber and have excellent resistance to heat and steam or water. This accounts for its use in gaskets and diaphragms for hot water and steam service. In addition, butyl rubber can be compounded to have low resilience properties and has found use in high damping mounts for engines, motors, and similar devices. Halobutyl rubbers can be blended with natural rubber, polychloroprene, and EPDM to greatly enhance their permeability resistance.

Nitrile Rubber. Nitrile rubbers are made by the emulsion copolymerization of acrylonitrile (9–50%) and butadiene (6) and designated NBR. The ratio of acrylonitrile (ACN) to butadiene has a direct effect on the properties on the nature of the polymers. As the ACN content increases, the oil resistance of the polymer increases (7). As the butadiene content increases, the low temperature properties of the polymer are improved (see ELASTOMERS, SYNTHETIC NITRILE RUBBER).

Nitrile rubber is much like SBR in its physical properties. It can be compounded for physical strength and abrasion resistance using traditional fillers such as carbon black, silica, and reinforcing clays. The primary benefit of the polymer is its oil and solvent resistance. At a medium ACN content of 34% the swell in IRM 903 oil at 70°C is typically 25–30%. Nitrile rubber processes on conventional rubber equipment and can be compression, transfer, or injection molded. It can also be extruded easily.

Nitrile rubber compounds have good abrasion and water resistance. They can have compression set properties as low as 25% with the selection of a proper cure system. The temperature range for the elastomers is from – 30 to 125°C. The compounds are also plasticized using polar ester plasticizers. The main dilemma is the selection of a heat-stable, nonfugitive plasticizer that also gives good low temperature properties.

In the United States, suppliers include Nippon Zeon, Bayer, Goodyear Chemical, Uniroyal, and several imported sources. There is an estimated 68,000 t yr used. Applications requiring good oil resistance and moderate heat aging are the principal uses of nitrile rubber compounds. The automotive industry has been the primary consumer of nitrile rubber goods. However, fluoroelastomers, silicone, epichlorohydrin, and acrylic elastomers have been replacing nitrile rubber for higher reliability and longer service life. Filler neck hoses, fuel lines, and numerous O-rings and seals in automotive applications are common uses. In addition numerous rolls in the printing industry are made from nitrile rubber (see PRINTING PROCESSES).

Hydrogenated Nitrile. Hydrogenation of nitrile rubber improves the aging properties of the oil-resistant backbone. The polymers are available with acrylonitrile contents from 17 to 50% and residual double-bond unsaturation from 2 to 20%. In addition to outstanding oil resistance and improved aging properties, the abrasion resistance of the polymers is very good. The basic polymer is quite polar giving inherent resistance to swelling in oils, fats, and gasoline, but can be affected by ketones, halogenated solvents and oils, and oxygenated solvents. They have a service range from – 45 to 149°C.

Two suppliers to the U.S. market are Bayer and Nippon Zeon. The estimated volume used in the United States is ~500–700 t yr. Hydrogenated nitrile rubber (HSN or HNBR) compounds compete with FKM, FVMQ, ECO, ACM, and high ACN content nitrile rubbers. Big applications include automotive timing belts, blowout preventors, drill pipe protectors, and numerous oil and fuel pump components.

Specialty Elastomers. Polychloroprene and polysulfide rubber were the first synthetic specialty elastomers discovered. Since their invention in the 1930s the total number of classes of synthetic rubbers has grown to almost 30. The following lists standard acronyms by the International Synthetic Rubber Producers (IISRP) and the American Society for Testing and Materials (ASTM) for several specialty elastomers.

Elastomer	Designation
acrylic elastomers	ACM, AEM, EEA
chlorinated polyethylene	CM
chloroprene	CR
chlorosulfonated polyethylene	CSM
epichlorohydrin	ECO
fluoroelastomers	FKM, FPM, FFKM
fluorosilicone	FVMQ
polysulfides	T
silicone	MQ
urethanes	AU, EU
vinyl acetate copolymers	EVA

Chloroprene Elastomers. Polychloroprene is a polymer of 2-chloro-1,3-butadiene. The elastomer is largely composed of the trans isomer. There are two basic polymer types: the W-type and the G-type. G-types are made by using a sulfur-modified process; W-types use no sulfur modification. As a result, G-types possess excellent processing and dynamic properties, and tend to be used in V-belts. However, they have poorer aging properties than W-types. The W-types tend to be used in applications requiring better aging, such as rolls and mechanical goods (see ELASTOMERS, SYNTHETIC POLYCHLOROPRENE).

Chloroprene elastomers have a balance of good properties. Essentially an isoprene elastomer with chlorine present in the backbone, the polymer exhibits excellent tensile strength and low hysteresis, much like natural rubber. Tensile properties up to 28 MPa (4000 psi) are possible with the proper reinforcing system. The polarity imparted by the chlorine, however, improves the oil and solvent resistance to a point approaching nitrile polymers. The polymer can be protected with paraphenylenediamines for good ozone resistance, and heat aging is also good. As a result, chloroprene elastomers are used in a wide variety of applications needing a balance of properties.

Principal uses include automotive V-belts, industrial and hydraulic hose, specialty roofing, heels and soles in footwear, wire covering, and a wide variety of coated fabric uses, eg, rafts. Chloroprene elastomers are also used extensively in adhesives (qv). It is estimated that about 77,000 t of chloroprene are used each year in the United States. The two main suppliers of chloroprene elastomers in the United States are DuPont and Bayer. In addition, Distiguil (France) sells polymers through the A. Schulman Company.

Chlorinated Polyethylene. Chlorinating polyethylene under pressure results in a polymer having a chlorine content varying from 25 to 42% . The polymer requires the incorporation of carbon black and minerals for achieving good physical properties. The polymers handle like conventional polymers and can be mixed and processed on conventional rubber equipment.

Chlorinated polyethylene (CPE) has excellent ozone, oil, and heat resistance. In addition chlorinated polyethylene has replaced chloroprene elastomers. CPE has a lower specific gravity than chloroprene compounds and produces compounds that are similar to CR in properties but with lower costs. In addition, due to high levels of chlorine in the polymer, the flame resistance of the compounds of CPE are high.

DuPont–Dow is the primary supplier of these polymers. There is an estimated 18,000 t of these elastomers used per year. The main uses of CPE are in construction, automotive, and electrical applications. These include power steering hose, electrical cords used in low voltage applications (extension cords, ignition wire), pond liners, and as a plastic modifier to improve impact modification.

Chlorosulfonated Polyethylene. This elastomer is made by the simultaneous chlorination and chlorosulfonation of polyethylene in an inert solvent. The resulting polymer is an ododess, colorless chip that is mixed and processed on conventional rubber equipment. The polymer typically contains 20–40% chlorine and 1% sulfur groups (see ELASTOMERS, SYNTHETIC CHLOROSULFONATED POLYETHYLENE) (8).

The polymer requires compounding with normal fillers to produce useful compounds. Chlorosulfonated polyethylene (CSM) excels in resistance to attack by oxygen, ozone, corrosive chemicals, and oil, and in addition has very good electrical properties. Electrical stability and resistance to corona and arc are good. The physical properties and abrasion resistance are also good. Light-colored goods made from CSM have excellent color-fastness. Due to the presence of chlorine atoms, this elastomer shows excellent flame resistance.

CSM is extensively used in construction and electrical applications. This includes roofing membranes, automotive ignition boots and wire, roll compounds, and in some automotive hoses requiring good heat and oil resistance, eg, air conditioning and power steering. It is also used in nuclear power plants because of its excellent resistance to radiation degradation.

It is estimated that 27,000 t yr of CSM have been commercially used in the United States. However, due to environmental problems in the manufacturing process, it has been necessary to develop a process that is much more expensive. As a result many companies using CSM are trying to replace the CSM with CPE or other elastomers. The result is a decline in the usage of this polymer. Chlorosulfonated polyethylene is sold under the trade name Hypalon (DuPont–Dow Company).

Acrylic Elastomers. Acrylic elastomers possess good oil and heat resistance. They are made by polymerizing monomeric acid esters of ethyl or butyl acrylate and methoxyethyl acrylate or ethoxyethyl acrylate. They can be polymerized in emulsion, suspension, or solution systems (9) (see ELASTOMERS, SYNTHETIC ACRYLIC ELASTOMERS).

The polyacrylates are inherently tacky and soft. They require compounding with carbon black or mineral fillers for achieving useful properties. Their main properties are a temperature performance range of -40–175°C. They have good resistance to oils, oxidation, ozone, aliphatic solvents, sunlight, weathering, and exhibit low permeability to gases. The tensile strength is usually 7–14 MPa (1000–2000 psi). Compounds can be fabricated with durometers from 40 to 90 Shore A.

Enichem and DuPont are two suppliers of acrylic elastomers. It is estimated there are ~9000 t yr of these elastomers used in the United States. The principal uses of acrylic elastomers are in automotive applications requiring both oil and heat resistance. These include transmission, valve stem, crankshaft, pinion, and oilpan gaskets and seals. In addition hose, tubing, rolls, and belts are made from acrylic polymers.

Epichlorohydrin. Commercial polyester elastomers include both the homopolymer and the copolymer of epichlorohydrin with ethylene oxide. The very polar chloromethyl groups create basic resistance to oil for these polymers, and they have been extensively used in fuel lines; however, the desire for lower fuel permeation is causing a search to be made for other polymers (10) (see ELASTOMERS, SYNTHETIC POLYETHERS).

Epichlorohydrin (ECO) has excellent resistance to fuel and oil swell. The ECOs show a volume swell of 35% at room temperature compared to 70% for a medium ACN–nitrile rubber in ASTM Reference Fuel C. The copolymer has a low temperature brittle point of -40°C and the homopolymer, -15°C. An interesting property of these elastomers is a stable dynamic performance over a wide temperature range; however, the electrical properties are only average.

The principal uses of ECO are in automotive applications, eg, for fuel line and fuel vapor recirculation. There are also some downhole seal applications where ECO is used. About 7000 t yr of epichlorohydrin are used in the United States. This is all supplied by the Zeon Corporation.

Fluoroelastomers. The fluoroelastomers were introduced to the rubber industry in the late 1950s by the DuPont Company. They were made by modification of Teflon polymers and designed to have excellent heat and chemical resistance, but remain elastomeric in nature. They were very expensive and have found use in limited applications. However, with the increasing demand in the automotive and industrial market for improved reliability and longer life, the elastomeric fluoroelastomers have made significant inroads into these applications (see ELASTOMERS, SYNTHETIC FLUOROCARBON ELASTOMERS).

Fluoroelastomers excel compared to all other elastomers in heat, chemical, flame, weathering, fuel, and ozone resistance. In addition oil, oxygen, and water resistance are very good. The fluoroelastomers, however, are attacked by amines and some highly polar solvents. The abrasion resistance and low temperature properties are adequate for most applications.

The four suppliers to the U.S. market are DuPont, 3M, Ausimont, and Daiken. It is estimated that 7000–8000 t yr are used in the United States and growing rapidly. In addition DuPont and Daiken sell what in essence are elastomeric Teflons under the respective trade names of Kalrez and Chemraz.

These have outstanding resistance to heat, chemicals, and hot fuel vapors and are only sold as fabricated parts.

Parts made from fluoroelastomers are used in applications that justify their high cost, usually where the maintenance and replacement costs are high enough to offset the initial cost of the part. These include automotive applications such as valve stem seals, fuel injector components, radiator, crankcase and transmission seals, and carburetor needle tips. Numerous seals and gaskets in the marine, oilfield, and chemical processing industries employ fluoroelastomers. In addition, many hoses in the automotive and chemical industry are made entirely of fluoroelastomer compounds or have a veneer of the fluoroelastomer as a barrier exposed to the harsh environment. Seals and gaskets in military applications and the binder for flares and missile applications are made with fluoroelastomers.

Polysulfides. Polysulfides are also called Thiokol's (T) after the company that originally made them. They were discovered in 1927 (see POLYMERS CONTAINING SULFUR, POLYSULFIDES).

Thiokol elastomers possess fairly low tensile and tear properties. However, they have excellent resistance to both aliphatic and aromatic solvents at room temperature and slightly elevated temperatures. The Thiokol division of Morton International Corporation is the supplier of polysulfide elastomers in the United States. It is estimated that 1360–1600 t are used annually in the United States. The primary use of polysulfide is in seals, gaskets, rolls, and diaphragms where solvent resistance and low permeability are useful.

Silicone. Silicone elastomers have a siloxy backbone with methyl, vinyl, and phenyl groups attached. The elastomers are designated by their chemical composition, as follows: methyl silicone (MQ), methyl vinyl silicone (VMQ), methyl phenyl silicone (PMQ), methyl phenyl vinyl silicone (PVMQ), and fluorovinyl methyl silicone (FVMQ).

Silicone elastomers possess outstanding resistance to heat aging. The Si–O–Si backbone imparts resistance to oxygen, ozone, uv, and to some polar fluids. However, the strength of these elastomers is usually just adequate. They have low abrasion resistance and tear strength (see SILICON COMPOUNDS, SILICONES).

Fluorosilicone. By fluorinating the silicone polymer molecule it is possible to improve the solvent, fuel, and oil resistance of this already heat-resistant class of elastomers. The resulting polymers are especially useful in select automotive seals and gaskets as well as military and downhole oilfield parts.

Fluorosilicones (FVMQ) have excellent low temperature flexibility properties coupled with good oil, fuel, and solvent resistance and excellent aging properties. The materials are compounded and reinforced with fine particle fillers, especially silica. The materials are mixed and processed on especially clean equipment and are peroxide-cured.

Dow Corning, General Electric, and Shinetsu Chemical Company are the principal suppliers of these elastomers in the United States. It is estimated that the volume used in the United States is between 2700 and 3600 t yr. The primary uses of these elastomers are in O-rings and shaft seals, as well as wire and cable, and electrical connectors.

Urethanes. Urethane elastomers are prepared by the reaction of an isocyanate molecule with a high molecular weight ester or ether molecule. The result is either an elastomeric rubber form or a liquid prepolymer that can be vulcanized with an amine or a hydroxyl molecule (see URETHANE POLYMERS).

Urethanes are processed as rubber-like elastomers, cast systems, or thermoplastic elastomers. The elastomer form is mixed and processed on conventional rubber mills and internal mixers, and can be compression, transfer, or injection molded. The liquid prepolymers are cast using automatic metered casting machines, and the thermoplastic pellets are processed like all thermoplastic materials on traditional plastic equipment. The unique property of the urethanes is ultrahigh abrasion resistance in moderately high Shore A (75–95) durometers. In addition, tear, tensile, and resistance to many oils is very high. The main deficiencies of the urethanes are their resistance to heat over 100°C and that shear and sliding abrasion tend to make the polymers soft and gummy.

There are many suppliers of urethane systems in the United States. The TSE Company supplies the rubber form; many companies supply the prepolymer forms. It is estimated that close to 68,000 t of these rubbers are used in the United States if all forms are included. Uses include sport wheels (roller blades, ski wheels); printing, paper, and steel processing rolls; gears; pump liners; appliance components; and solid industrial tires.

Vinyl Acetate–Ethylene Copolymers. In these random copolymers, the ratio of ethylene to vinyl acetate (EVA) is varied from 30–60%. As the vinyl acetate content increases, the oil and heat resistance increases. With higher ethylene content the physical strength, tensile, and tear increases. The polymers are cured with peroxide. The main properties of these elastomers include heat resistance, moderate oil and solvent resistance, low compression set, good weather resistance, high damping, excellent ozone resistance, and they can be easily colored (see VINYL POLYMERS, POLY(VINYL ACETATE)).

Bayer, USI, and DuPont are U.S. suppliers. There is an estimated 2300–4000 t yr of EVA used in the United States. Uses include automotive electrical system components, eg, ignition cable, footwear, weatherstripping, and miscellaneous O-rings and seals for industrial applications.

Reclaim Rubber. The process of reclaiming rubber by chemical digestion has been in use since the late 1800s. Early processes involved the treatment of the rubber cord mix with acid. Acids attack cotton, rayon, and nylon. The acid treatment was used to remove the reinforcing components. The rubber was ground and treated. It was then washed and the rubber steam devulcanized and formed to shape where it could then be reincorporated into virgin rubber compounds. An improvement on this technique was the use of alkali digestion (see RECYCLING, RUBBER).

The principal benefit of using reclaim is its lower cost compared to virgin rubbers. Reclaim typically sells for 20–30% of the cost of its none-reclaimed counterparts. Reclaimed rubber also imparts some desirable improvements in processing; it has much lower nerve than virgin polymers. As a result, compounds containing reclaim have much lower die swell and extrusion rates. It also increases calender rates and improves flow and mold filling.

One of the shortcomings of reclaim is that it lowers the green strength and tensile strength of the compounds in which it is used. Reclaim is in fact a mixture of rubber, carbon black, oil, zinc oxide, stearic acid, and other compounding ingredients used in the original compounds. The levels of reclaim used are generally 15 to 45 phr, about half of which is rubber and the other half compounding ingredients. In some low performance requirement rubber products, eg, mats, much higher levels can be used (11–15).

Steam digestion of silicone rubber produces a silicone reclaim, used to reduce the cost of silicone compounds. It has been widely used by silicone automobile ignition systems. The resulting compounds have excellent aging characteristics and still possess outstanding electrical properties (16).

Reclaim is used in many tire applications: carcass and innerliner compounds, sidewalls, chafers, and rubber used in bead components. The introduction of radial tires changed the property needs for compounds used in tire manufacturing and reduced the volume of reclaim. Butyl reclaim made from tubes is used in the innerliner of both bias and radial tire compounds and some reclaim is used in low speed bias tires. There is a fairly large volume of some high quality rubbers such as natural rubber and butyl being reclaimed and used in some solid tires and other mechanical goods to extend the compound without having an adverse effect on the product.

Ground rubber has replaced reclaim in many tire applications including treads, carcasses, innerliners, subreads, and bead components. It is also used in mechanical goods, footwear, solid tires, mats, and retreads. Cryogenically and wet-ground rubbers which are of smaller particle size can be used at moderately high levels and still allow compounds to be processible.

Vulcanization

Vulcanization is a chemical process for improving an elastomer compound's performance. However, in most cases not all of the desired properties reach their optimum levels simultaneously. One of the rubber compounder's key responsibilities is to achieve a balance of the most important property requirements by the proper selection of cure system (chemical) and time–temperature cure cycle (physical).

Frequently, the curing equipment available, ie, presses, autoclaves, LCM lines, etc, do not allow the curing conditions to be varied as desired, so the compounder must design a cure system compatible with the existing equipment while also meeting the compound performance requirements.

Measuring Vulcanization. The formation of a three-dimensional structure during vulcanization increases the stiffness (modulus) of the compound. Therefore, following the modulus increase versus cure time provides a continuous picture of the vulcanization process. Oscillating disk rheometers provide a useful method to do this (17). In this test, a preweighed sample of uncured rubber is placed into a preheated cavity containing a conical rotor. The cavity is closed and the rotor is set to oscillate within the rubber sample. As vulcanization proceeds, the compound's resistance to rotor movement increases and this resistance is followed as a function of time, thereby generating a continuous profile of cure behavior. These cure curves,

shown in Figure 2, provide the compounder with information on scorch safety or cure induction time, cure rate, state of cure or stiffness developed, optimum cure time or time when modulus no longer increases significantly, and reversion or the tendency of the compound to degrade upon overcure.

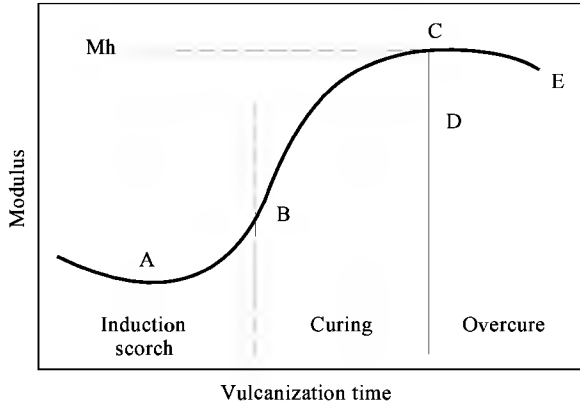


Fig. 2. Cure curve from oscillating disk rheometer where A represents scorch safety; B, cure rate; C, state of cure; D, optimum cure time; and E, reversion.

Effect on Elastomer Physical Properties. Sulfur vulcanization is a complex reaction during which both the number (density) and type (structure) of cross-links are continuously changing as the reaction proceeds. The chemical structures formed at any given time during cure may favor one set of properties such as tear strength while not being optimum for others such as hysteresis and set. A cure system designed to minimize these trade-offs is always a goal of the compounder.

The effect of both cross-link number and cross-link length (S_x) on the most common physical properties has been summarized (18) (Table 3). These trends can be explained as follows. Modulus or stiffness increases as the number of cross-links increases because the 3-D structure which forms becomes more resistant to deformation under load thereby requiring greater force to reach a given elongation. With overcure, cross-link structure either degrades and modulus drops as in natural rubber, or it continues to build, resulting in marching modulus as in SBR and polybutadiene.

Table 3. Effect of Increasing Cross-Link Number and Length^a

Property	Number	Length
tensile strength	>, then <	>
elongation	<	>
modulus	>	<
flex life	<	>
compression set	<	>
hysteresis	<	>
resilience	>	<
thermal stability		<
oxidative stability		<
rolling resistance	<	>
tear strength	>, then <	>
abrasion	>	>

^a Ref. 18.

Tensile and tear strengths and fatigue life frequently pass through a maximum as cure progresses. These ultimate properties are highly dependent on the presence of flaws in the material. As cross-linking proceeds and modulus builds, any applied load acting on a flaw (from poor dispersion, small nicks or tears, localized overcure, etc) concentrates the stresses at the flaw and provides a mechanism for tensile failure and tear propagation or poor fatigue life (Fig. 3).

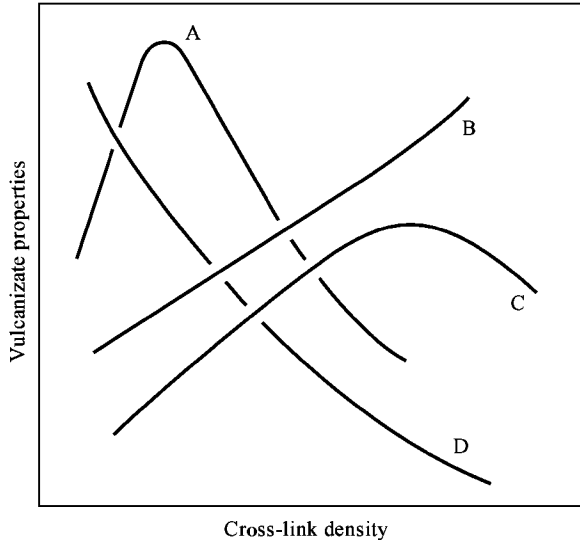


Fig. 3. Effect of cross-link density where A represents tear strength, fatigue life, and toughness; B, elastic recovery and stiffness; C, strength; and D, hysteresis, permanent set, and friction coefficient.

Permanent set and low hysteresis properties depend on minimizing the viscous or plastic component of modulus. Because cross-linking increases elasticity, a high state of cure typically provides the best compression set and heat buildup properties.

Heat resistance is influenced by both the type and extent of cure. The greater the strength of the chemical bonds in the cross-link, the better is the compound's heat resistance. Peroxide cure systems, which result in carbon–carbon bonds, result in a range of sulfur cross-links varying from 1 to >30 sulfur atoms per cross-link, and heat resistance improves as the number of more thermally stable short cross-links predominates. This is an important factor in designing the desired cure system.

Some properties are not significantly dependent on cross-linking and remain nearly invariant as cure progresses. These include thermal conductivity, electrical properties, and low temperature brittleness.

Another cure system consideration is the compound scorch behavior. Prior to vulcanization, rubber is plastic-like and can be processed into desired shapes such as tires, hoses, belts, or other articles. The time available to accomplish this processing depends largely on the cure system and is referred to as the scorch time. If a compound cures prematurely during the processing step, it usually becomes useless scrap. Therefore, a key requirement of the vulcanization step is to minimize premature vulcanization or scorch (Fig. 4).

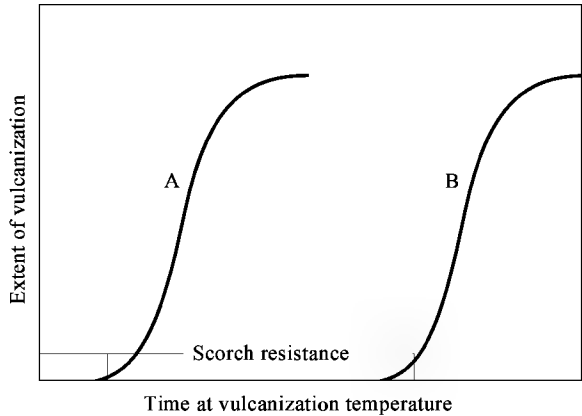


Fig. 4. Effect of heat history on processing (scorch) safety where A shows the time after processing and storage and B shows vulcanization without processing heat history. The scorch resistance represents delay time.

VULCANIZING AGENTS

Sulfur and Sulfur Donors. Sulfur is the most common vulcanizing agent for the widely used diene-containing elastomers, such as natural rubber, SBR, and polybutadiene. Rubbermakers' sulfur is a rhombic form existing as a cyclic or eight-member ring structure. Insoluble sulfur is an amorphous polymeric form with molecular weights of 100,000–300,000. Because this material is insoluble in rubber it resists migration to the surface prior to cure and this bloom-free attribute contributes to maintaining better building tack and better interply adhesion (19,20).

Sulfur-containing chemicals such as dimorpholinyl disulfide (DTDM) and tetraethylthiuram disulfide (TMTD) are not only effective accelerators, but they can also be used as sulfur donors. As such, they are effective in controlling sulfur cross-link length to form primarily mono- and disulfide cross-links. These short cross-links are more thermally stable than conventional sulfur curing and thereby provide better heat and set resistance.

Nonsulfur Vulcanizing Agents. Many high performance specialty elastomers do not contain diene moieties in their molecular structure and therefore cannot be sulfur-cured. These elastomers require cross-linking agents capable of reacting with the specific functional group(s) contained by the specific elastomer. Some common nonsulfur curatives include peroxides, difunctional resins, and metal oxides.

Peroxides. Peroxides are probably the most common materials used after sulfur because of their ability to cross-link a variety of diene- and nondiene-containing elastomers, and their ability to produce thermally stable carbon–carbon cross-links. Carbon–carbon bonds are inherently stronger than the carbon–sulfur bonds developed with sulfur vulcanization (21).

Peroxides decompose when heated to produce active free radicals which in turn react with the rubber to produce cross-links. The rate of peroxide cure is controlled by temperature and selection of the specific peroxide, based on half-life considerations (see INITIATORS, FREE RADICAL; PEROXY COMPOUNDS, ORGANIC). Although some chemicals, such as bismaleimides, triallyl isocyanurate, and diallyl phthalate, act as coagents in peroxide cures, they are not vulcanization accelerators. Instead they act to improve cross-link efficiency (cross-linking vs scission), but not rate of cross-link formation.

Metal Oxides. Halogen-containing elastomers such as polychloroprene and chlorosulfonated polyethylene are cross-linked by their reaction with metal oxides, typically zinc oxide. The metal oxide reacts with halogen groups in the polymer to produce an active intermediate which then reacts further to produce carbon–carbon cross-links. Zinc chloride is liberated as a by-product and it serves as an autocatalyst for this reaction. Magnesium oxide is typically used with ZnCl₂ to control the cure rate and minimize premature cross-linking (scorch).

Amine Cross-Linking. Two commercially important, high performance elastomers which are not normally sulfur-cured are the fluoroelastomers (FKM) and the polyacrylates (ACM). Polyacrylates typically contain a small percent of a reactive monomer designed to react with amine curatives such as hexamethylene-diamine carbamate (Diak #1). Because the type and level of reactive monomer varies with ACM type, it is important to match the curative type to the particular ACM in question. Sulfur and sulfur-bearing materials can be used as cure retarders; they also serve as age resistors (22). Fluoroelastomer cure systems typically utilize amines as the primary cross-linking agent and metal oxides as acid acceptors.

Other Difunctional Compounds. Other examples of difunctional compounds are epoxy resins for nitrile rubber, quinone dioximes for butyl rubber, and phenolic resins for butyl rubber, EPDM, and some EPDM-based thermoplastic elastomers (TPEs). The cross-linking mechanism involves a condensation reaction between each of the two curative difunctional groups with the polymer chain to form a three-dimensional network. Because only carbon–carbon linkages result, thermal stability and permanent set resistance are excellent, often approaching those of peroxide curing. A summary of the advantages and disadvantages of each of the various cross-linking agents appears in Table 4.

Table 4. Vulcanizing Agents

Class	Elastomer	Advantage	Disadvantage
accelerated sulfur and sulfur donor	diene ^a	versatile	heat resistance, set properties
peroxides	most elastomers, especially saturated ^b	excellent heat set resistance	control of cure rate, poor fatigue resistance
resins	primarily butyl rubber	heat resistance, stable modulus	slow cure
metal oxides	halogenated elastomers	best with CR, CSM	water resistance

^a For example, NR, SBR, BR, and EPDM.
^b For example, EPM.

RUBBER VULCANIZATION CHEMICALS

Accelerators. During sulfur vulcanization of rubber, accelerators serve to control time to onset of vulcanization, rate of vulcanization, and number and type of sulfur cross-links that form. These factors in turn play a significant role in determining the performance properties of the vulcanizate. There are seven principal classes of accelerators and several miscellaneous products that do not fit into these classes. In addition, many proprietary blends of several accelerators are sold which are designed as cure packages for a specific applications. Choosing the best cure system is a responsibility of the rubber chemist and requires extensive knowledge of each accelerator type and its applicability in each elastomer. Table 5 shows a rule of thumb comparison of the scorch cure rate attributes for the five most widely used classes of accelerators used in the high volume diene-based elastomers.

Table 5. Accelerated Sulfur Vulcanization

Accelerator type	Scorch safety	Cure rate	Cross-link length
none		very slow	very long
guanidines	moderate	moderate	medium-long
mercaptobenzothiazoles	moderate	moderate	medium
sulfenamides	long	fast	short-medium
thiurams	short	very fast	short
dithiocarbamates	least	very fast	short

Although differences in accelerator response do occur from one rubber type to another because of inherent differences in their molecular structures, the guidelines in Figure 5 serve as an overview of the principal accelerators and their performance (see RUBBER CHEMICALS). It is common practice in the rubber industry for a compounder to use combinations of several accelerators in developing a cure system. Typically these cure systems are comprised of a primary accelerator and one or more secondary types. Primary accelerators are generally the thiazole and sulfenamide classes; the secondary types (kickers) are the thiurams, dithiocarbamates, guanidines, and to a much lesser extent, certain amines and the dialkylphosphorodithioates (20).

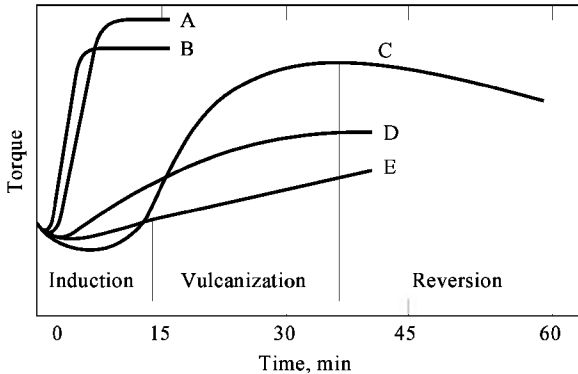


Fig. 5. Cure characteristics of accelerators: A, thiuram; B, dithiocarbamate; C, sulfenamide; D, thiazole; and E, guanidine. The induction period represents scorch time.

As a general rule the sulfenamides exhibit faster cure rate than the thiazoles. If secondary accelerators are used, dithiocarbamates are scorchiest and give the fastest cure followed by the thiurams, then the guanidines. Figure 6 summarizes these comparisons to show a series of natural rubber (NR) recipes using either a thiazole (MBTS) or sulfenamide (TBBS) primary accelerator in combination with the various secondary accelerators (21). In this study, the initial primary accelerator levels were selected to produce nearly equivalent modulus or state of cure in the NR.

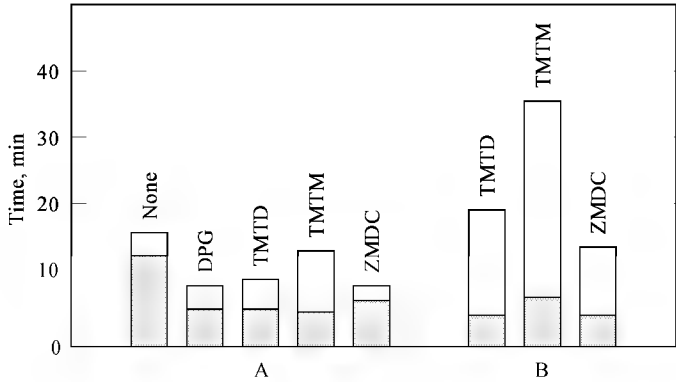


Fig. 6. Comparison of secondary accelerators where □ represents Mooney scorch at 121°C and ▒, optimum (t90) at 153°C: A, the primary accelerator is 0.75 phr MBTS and the secondary accelerator is at 0.25 phr; B, 0.45 phr TBBS is the primary accelerator and there is 0.2 phr secondary accelerator.

Activators. Activators are chemicals that increase the rate of vulcanization by reacting first with the accelerators to form rubber soluble complexes. These complexes then react with the sulfur to achieve vulcanization. The most common activators are combinations of zinc oxide and stearic acid. Other metal oxides have been used for specific purposes, ie, lead, cadmium, etc, and other fatty acids used include lauric, oleic, and propionic acids. Soluble zinc salts of fatty acid such as zinc 2-ethylhexanoate are also used, and these rubber-soluble activators are effective in natural rubber to produce low set, low creep compounds used in load-bearing applications. Weak amines and amino alcohols have also been used as activators in combination with the metal oxides. Natural rubber usually contains sufficient levels of naturally occurring fatty acids to solubilize the zinc salt. However, if these fatty acids are first extracted by acetone, the resultant "clean" natural rubber exhibits a much lower state of cure. Therefore, to ensure consistent cure rate, fatty acids are usually added. Synthetic rubbers, especially the solution polymers, do not contain fatty acids and require their addition to the cure system. Sulfenamide accelerators generally require less fatty acid because they release an amine during the vulcanization process which acts to solubilize the zinc. Guanidines and similar amine accelerators also serve to both activate and accelerate vulcanization.

A study of the effect of stearic acid and zinc oxide on a sulfonamide-accelerated, sulfur-cured natural rubber compound dramatically showed the need for both zinc and fatty acid activators (Fig. 7) (21).

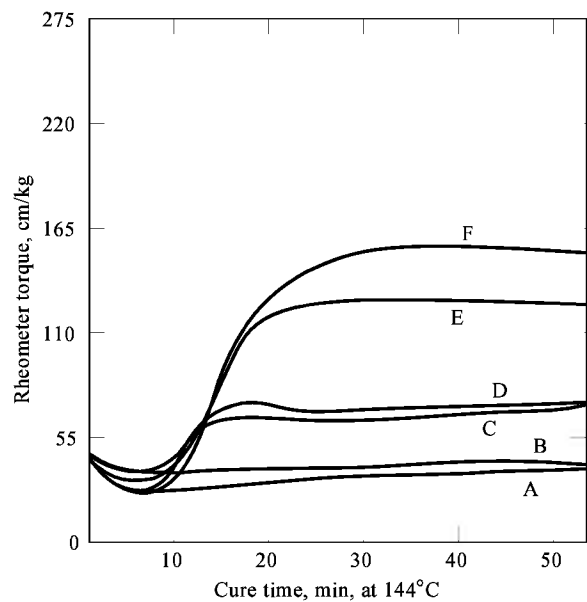


Fig. 7. Effect of activators on cure rate where A is 2.5 phr sulfur; B, sulfur + stearic acid (2 phr) + zinc oxide (5 phr); C, sulfur + TBBS (0.6 phr); D, sulfur + TBBS + stearic acid; E, sulfur + TBBS + zinc oxide; and F, sulfur + TBBS + stearic acid + zinc oxide. To convert cm kg to in. lb, divide by 5.5.

The role of activators in the mechanism of vulcanization is as follows. The soluble zinc salt forms a complex with the accelerator and sulfur. This complex then reacts with a diene elastomer to form a rubber–sulfur–accelerator cross-link cursor while also liberating the zinc ion. The final step involves completion of the sulfur cross-link to another rubber diene segment (18).

Retarders. The purpose of vulcanization retarders is to delay the initial onset of cure in order to guarantee sufficient time to process the unvulcanized rubber. Three main classes of materials are used commercially, including organic acids and anhydrides, cyclohexylthiophthalimide (Santogard PVI or CTP), and a sulfenamide material (Vulkalent E).

Examples of organic acid retarders include phthalic anhydride and benzoic and salicylic acids. These materials are thought to function by reaction with basic components present as accelerator fragments, from other basic compounding ingredients, and from impurities. These basic moieties, which would normally serve to accelerate vulcanization and produce a higher state of cure, are neutralized by the acid retarders. Therefore, they are effective in delaying the initial onset of cure. However, they also retard cure rate and in many cases detract from the final performance properties. This retardation of cure rate and loss in performance is a high price to pay for improved scorch safety (17).

The thiophthalimide (CTP) and sulfenamide classes of retarders differ from the organic acid types by their ability to retard scorch (onset of vulcanization) without significantly affecting cure rate or performance properties. Much has been published on the mechanism of CTP retardation. It functions particularly well with sulfenamide-accelerated diene polymers, typically those used in the tire industry. During the initial stages of vulcanization, sulfenamides decompose to form mercaptobenzothiazole (MBT) and an amine. The MBT formed reacts with additional sulfenamide to complete the vulcanization process. If the MBT initially formed is removed as soon as it forms, vulcanization does not occur. It is the role of CTP to remove MBT as it forms. The retardation effect is linear with CTP concentration and allows for excellent control of scorch behavior.

Another commercially available retarder for sulfur vulcanization is based on an aromatic sulfenamide. Like CTP, this product is most effective in sulfenamide cure systems, but it also works well in thiazole systems. Performance properties are generally not affected except for a slight modulus increase. In some cases this feature allows for the use of lower levels of accelerator to achieve the desired modulus with the added potential benefits of further scorch delay and lower cost cure system (23).

CURE SYSTEM DESIGN

Curatives and accelerators are combined to achieve desired performance properties through cure system design.

There are three generally recognized classifications for sulfur vulcanization: conventional, efficient (EV) cures, and semiefficient (semi-EV) cures. These differ primarily in the type of sulfur cross-links that form, which in turn significantly influences the vulcanizate properties (Fig. 8) (21). The term efficient refers to the number of sulfur atoms per cross-link; an efficiency factor (E) has been proposed (20).

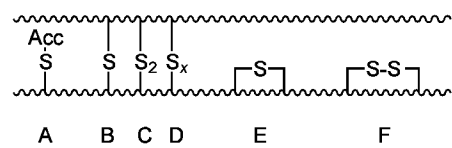


Fig. 8. Sulfur-based cure system designs where conventional systems are polysulfidic, EV systems are mono- disulfidic, and semi-EV systems are clean polysulfidic. A shows pendent sulfide groups terminated by accelerator; B, monosulfide cross-links; C, disulfide cross-links; D, polysulfide cross-links ($x > 2$); E, cyclic monosulfides; and F, cyclic disulfides.

Conventional cure systems use relatively high levels (2.5 + phr) of sulfur combined with lower levels of accelerator(s). These typically provide high initial physical properties, tensile and tear strengths, and good initial fatigue life, but with a greater tendency to lose these properties after heat aging.

In contrast, the EV cure systems employ much lower levels of free sulfur (0.1–1.0 phr) or they use sulfur donors such as TMTD or D'TDM combined with higher accelerator levels. The short mono- and disulfide cross-links that form often do not exhibit the excellent physical properties of the conventional systems but they do retain their properties much better after aging.

Semi-EV cures represent a compromise between conventional and EV cures. Although semi-EV cures do yield polysulfide cross-links, they tend to minimize formation of inefficient moieties such as sulfur bridging with itself, accelerator-terminated sulfur linkages, etc. This cleaner usage of sulfur is the reason for their compromise properties between conventional and EV cures.

The reason for the greater loss in properties with conventional cure systems can be understood by examining cross-link bond strengths. Polysulfide bond strengths, from conventional cures, are significantly lower than bond strengths from the shorter cross-links obtained with EV cure systems. The relatively unstable S_x bonds break and rearrange to form mono- and disulfide linkages plus noncross-linking cyclic and or accelerator-terminated

fragments. These chemical reactions are observed as changes in properties with increased aging severity.

However, conventional systems in natural rubber do provide better flex life than EV cures, and this is one of the limitations of EV curing. The short monosulfide bonds are less able to rearrange to relieve localized stresses which build during flexing, whereas the longer S_x bonds can. This ability for stress relief is thought to be the mechanism for the superior flex life of conventional cures.

If natural rubber compounds are subjected to thermal aging plus fatigue, the conventional systems perform no better than EV systems. The compromise obtained by using semi-EV systems involves the balance between heat aging and flex life.

Examples of Cure Systems in NR, SBR, and Nitrile Rubber. Table 6 offers examples of recipes for conventional, semi-EV, and EV cure systems in a simple, carbon black-filled natural rubber compound cured to optimum (t90) cure. The distribution of cross-links obtained is found in Figure 9 (24).

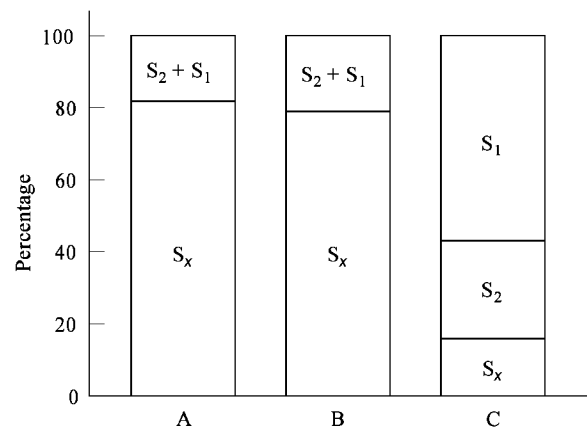


Fig. 9. Distribution of sulfur cross-links in a carbon black-filled NR, where A is the conventional, B the semi-EV, and C the EV system.

Table 6. Recipes for Cure Systems^a

Ingredients	Conventional	Semi-EV	EV
SMR-5	100	100	100
N330 black	50	50	50
ZnO	3	3	3
stearic acid	1	1	1
6 ppd wax	1	1	1
sulfur	2.5	1.5	0.15
CBS	0.6	1	2
TMTD			2.3

^a In carbon black-filled natural rubber compound.

Clearly, the EV system exhibits the greatest number of thermally stable S_1 and S_2 cross-links. As expected, the conventional and semi-EV look similar in cross-link distribution, but again, the semi-EV system has fewer extraneous moieties. Physical properties of these compounds are summarized in Figure 10.

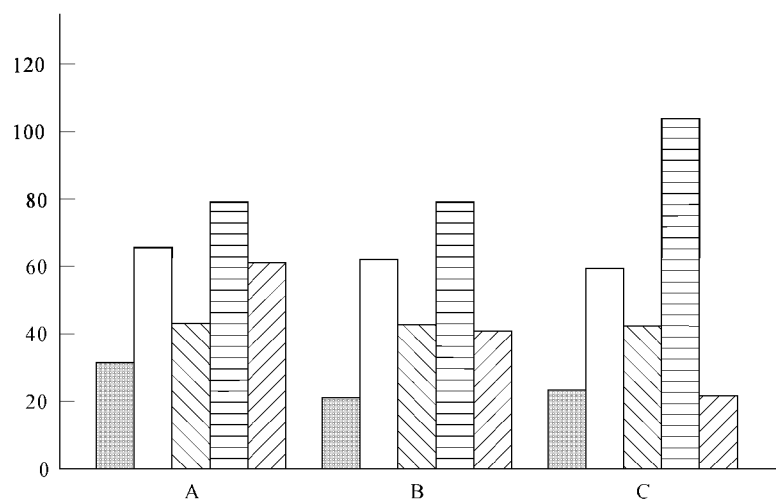


Fig. 10. Physical properties of compounds where A is the conventional, B the semi-EV, and C the EV system: ▨ represents tensile strength in MPa; □, hardness, Shore A; ▩ elongation, %; ▤, $\tan \delta$; and ▧, fatigue life, kc. To convert MPa to psi, multiply by 145.

All three systems give similar tensile strength, elongation, and hardness properties. Hysteresis (heat buildup) measured by $\tan \delta$ shows an advantage for conventional and semi-EV systems and unaged fatigue follows the expected pattern.

Figure 11 summarizes properties obtained after heat aging these compounds at 70°C. Both EV and semi-EV systems are superior in tensile and hardness retention, and the EV system is inferior in aged flex life. Loss of fatigue is greatest for the conventional cure. These observations illustrate the importance of achieving the necessary cross-link structures for an optimum balance of properties. These illustrations are for compounds subjected to ideal or t90 cure conditions at a relatively mild 140°C temperature. Extremely high cure temperatures, such as those used in injection molding, can be expected to alter these results, but not the overall trends.

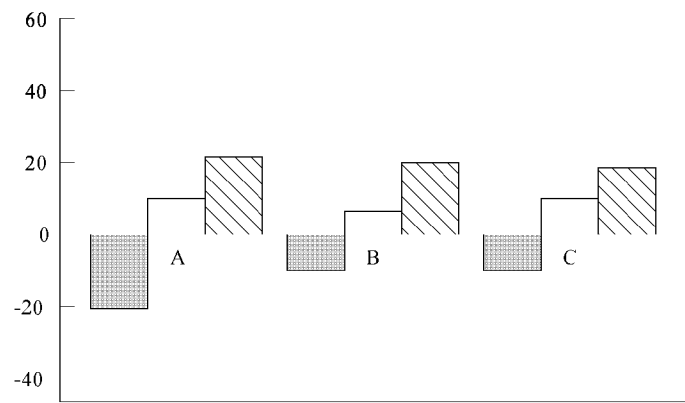


Fig. 11. Aging properties of cured natural rubber for 70 hours at 70°C. A is the conventional, B the semi-EV, and C the EV system where ▨ shows tensile loss, %; □, flex fatigue life, kc; and ▤, hardness change.

In a study of how cross-link structure of a conventionally cured compound changes as a function of cure time, it was seen that total cross-links decrease rapidly past optimum cure (reversion) and that most of the cross-links lost are the S_x types as cure progresses (22). Cure temperature affects cross-link density; there is a dramatic decrease in cross-link density as cure temperature increases. This is a serious concern for the compounder charged with increasing productivity by going to higher temperature, shorter cure cycles. Although plant efficiencies might improve, compound performance is likely to suffer.

SBR and Nitrile Rubber. EV and semi-EV cure systems can also be used effectively in other diene type elastomers to improve aging properties. SBR in particular can take advantage of EV curing because SBR develops a higher level of polysulfide cross-links than NR with EV cure systems thereby retaining good fatigue properties. Also, the average number of sulfur atoms per cross-link is significantly shorter in SBR than in natural rubber, and this results in a somewhat more thermally stable, conventionally cured network.

Although SBR does not exhibit the cure system sensitivity of natural rubber, use of EV and semi-EV systems is still advantageous as shown in Figure 12 (21). Superior heat aging, measured by elongation retention and resistance to stiffening upon overcure, is apparent for the EV cure. Interestingly, fatigue life before and after aging also favors the EV system.

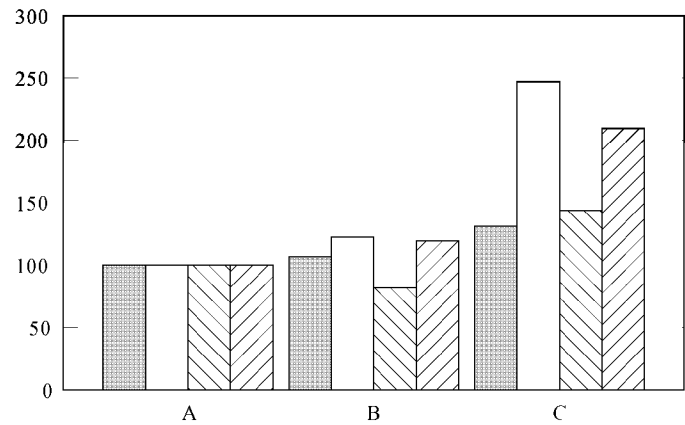


Fig. 12. SBR cure systems (A - conventional, B - semi-EV, C - EV) relative to conventional cure at 100 where ▨ shows aged elongation retained; □, overcure hardening retained; ▤, initial fatigue life; and ▥, aged fatigue life.

The effectiveness of EV curing in nitrile rubber (NBR) has also been demonstrated, but there is the potential for unsightly bloom which can occur with high sulfur or accelerator levels. The improved heat resistance obtained with EV cures is useful for severe areas such as under the hood in automotive systems and downhole oilwell applications. In general, NBR can be successfully cured with accelerator systems similar to those used in SBR. Care must be taken, however, to ensure proper sulfur dispersion. Elemental sulfur is incompatible with NBR and this can lead to poor sulfur dispersion and inconsistent performance. Magnesium carbonate treated sulfur (MC sulfur) is usually used to facilitate dispersion. Use of sulfur donors such as DTDM can eliminate the dispersion problem and provide superior heat aging. Strategies for developing NBR cure systems for specific attributes were published by B. F. Goodrich in the 1960s and are shown in Figure 13. Although these require further refinement for a specific application, the strategies have withstood the test of time and are still useful (25).

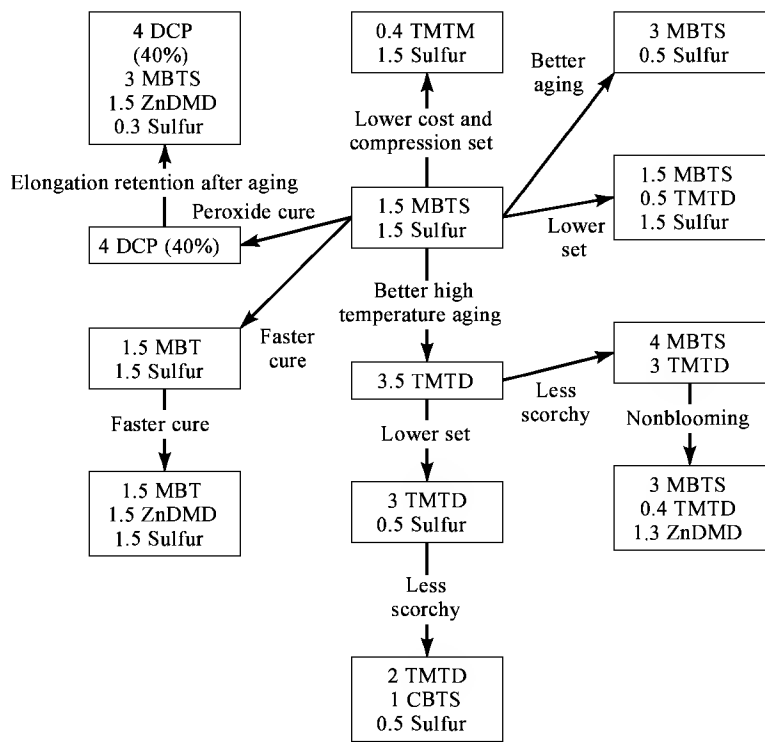


Fig. 13. Relationship of nitrile rubber cure systems where DCP is dicumyl peroxide; MBTS, benzothiazyl disulfide; ZnDMD, zinc dimethyldithiocarbamate; MBT, mercaptobenzothiazole; TMTM, tetramethylthiuram monosulfide; TMTD, tetramethylthiuram disulfide; and CBTS, *N*-cyclohexyl-2-benzothiazole sulfenamide.

Cure Systems of Butyl Rubber and EPDM. Nonhalogenated butyl rubber is a copolymer of isobutylene with a small percentage of isoprene which provides cross-linking sites. Because the level of unsaturation is low relative to natural rubber or SBR, cure system design generally requires higher levels of fast accelerators such as the dithiocarbamates. Examples of typical butyl rubber cure systems, their attributes, and principal applications have been reviewed (26). Use of conventional and semi-EV techniques can be used in butyl rubber as shown in Table 7 (21).

Table 7. Conventional and Semi-EV Cure Systems for Butyl Rubber^a

Ingredient	Semi-EV	Conventional
sulfur	0.5	2.0
TMTD	1.0	1.0
MBT		0.5
DTDM	1.2	
TBBS	0.5	
<i>Processing properties^b</i>		
Mooney scorch at 121°C		
t15, min	36.2	18.5
rheometer at 165°C		
t90, min	21.0	21.8
<i>Physical properties</i>		
cured to t90 at 165°C		
Hardness, Shore A	68	68
300° modulus, MPa	5.5	7.1
ultimate tensile, MPa	11.0	11.3
ultimate elongation, %	600	510
heat age resistance, 70 h at 121°C		
retained tensile, %	76	57
compression set, %	56	81

^a Masterbatch butyl 218-100, GPF black-70, paraffinic oil-25, zinc oxide-5, and promoter-1.

^b t90 optimal; t5, less than optimal.

EPDM is a terpolymer of ethylene, propylene, and a small amount (<10%) of an unsaturated diene third monomer to provide a cure site. Unlike the elastomers previously discussed, the unsaturation in EPDM is not in the main chain, but it is pendent to the chain. Peroxide cure gives superior aging resistance and low compression set.

A comparison of a sulfur-based cure to two different peroxides in EPDM is shown in Table 8 (27). Initial properties for these three compounds are reasonably close. However, after air aging, the advantages of peroxide curing are apparent. Most dramatic is the improved compression set obtained by the highly stable C–C cross-links developed with peroxides. Other advantages include ability to achieve high transparency or nondiscoloring and very low creep or stress relaxation. Disadvantages limiting the use of peroxides include poor hot tensile and tear strengths, slow cure rate, a tendency for scorch problems, and somewhat more difficult storage and handling problems in the plant.

Table 8. Peroxide vs Sulfur-Cured EPDM

Product	Sulfur	Di-cup	Vul-cup
Nordel ^a 1040	100	100	100
HAF black, phr	50	50	50
ZnO, phr	5.0	5.0	5.0

stearic acid, phr	1.0		
sulfur, phr	1.5		
TMTM, ^b phr	1.5		
MBT, ^c phr	0.5		
40KE, phr		6.5	4.1
TMQ, ^d phr	0.5	0.5	0.5
cure temperature, °C	160	171	177
cure time, min	20	20	20
<i>Original properties</i>			
modulus, MPa ^e			
100° o	2.2	1.8	1.9
200° o	6.2	5	5.4
tensile strength, MPa ^e	17.9	17.2	18.1
elongation, % o	400	375	375
hardness, points	68	62	60
Air-aged, 70 h at 150°C			
modulus, MPa ^e			
100° o	5.38	1.7	2.1
200° o	12.8	4.7	5.2
tensile strength, MPa ^e	15	17.4	16.8
elongation, % o	220	400	350
hardness, points	78	58	60
compression set, % o	77	21	19

^a E. I. du Pont de Nemours & Co., Inc.

^b TMTM = tetramethylthiuram monosulfide.

^c MBT = 2-mercaptobenzothiazole.

^d TMQ = polymerized trimethyldihydroquinoline.

^e To convert MPa to psi, multiply by 145.

General Strategy (Statistical Techniques). It is apparent from the above discussions that achieving the proper balance of properties by cure system design can be a tedious process. For this reason, compounders often use statistically designed experiments to optimize curative levels (28). By using properly designed experiments, ie, balanced or orthogonal arrays of curative levels, regression analyses can be performed which unequivocally quantify the effect of each ingredient, and combination of ingredients, on the properties of interest. From this, an optimum balance of properties is obtained much more efficiently than the old-fashioned "vary one ingredient at a time" approach. It is encouraged that this technique be used for solving problems in cure system design or any problem involving optimizing combinations of materials.

EFFECT OF OTHER COMPOUNDING INGREDIENTS ON VULCANIZATION

Other ingredients besides the elastomer and the cure system itself influence cure and scorch behavior. Usually the effect of a material on cure is pH-dependent. Ingredients which are basic in nature tend to accelerate the rate of both scorch and cure, whereas acidic materials exhibit the opposite effect.

Antidegradants. Amine-type antioxidants (qv) or antiozonants (qv) such as the phenylenediamines (ppd) can significantly decrease scorch time. This is particularly true in metal oxide curing of polychloroprene or in cases where the ppd had suffered premature degradation prior to cure. In one experiment the effect of ppd assay was correlated to scorch safety. As the ppd degrades liberate free amine, scorch time decreases and cure rate is faster. The degradation products apparently serve to activate the cure, since both the induction time, *t*₂, and cure time, *t*₀₀, decrease with decreasing ppd assay. However, the effect on unaged properties is minimal.

In contrast, antioxidants can have an opposite effect when peroxide curing. Because peroxide cross-linking involves a free-radical mechanism, and antioxidants are designed to scavenge free radicals, it is obvious that peroxide efficiency can be compromised by the addition of antioxidants. Thus the decomposition products of the ppds were acting as accelerators (29).

Fillers. Materials used as fillers (qv) in rubber can also be classified as acidic, basic, or neutral. Furnace blacks, ie, HAF, FEF, or SRF, are somewhat basic. As such, they can have an activating effect on sulfur cure rates. Furthermore, carbon blacks have been found to promote formation of mono disulfide cross-links thereby helping minimize reversion and enhance aging properties.

Channel blacks such as the old EPC and MPC grades are acidic to neutral and can vary from having little effect to having a slight retardation in cure rate. The pH-neutral large-particle size MT thermal black generally has little effect on cure rate.

Nonblack fillers such as the precipitated silicas can reduce both rate and state of cure. The mechanism appears to be one of a competitive reaction between rubber and filler for the zinc oxide activator. Use of materials such as diethylene glycol or triethanolamine prevents this competition thereby maintaining the desired cure characteristics. Neutral fillers such as calcium carbonate (whiting) and clays have little or no effect on the cure properties.

Process Oils, Plasticizers. Petroleum-based rubber process oils generally contain a mixture of paraffinic, naphthenic, and aromatic components. These oils vary in composition from grade to grade, but most contain some unsaturated moieties and this unsaturation can compete with the polymer for curatives. Therefore, state of cure can be decreased. This is not easily detected because oil softens the compound which masks the loss of state of cure.

Plasticizers (qv) can range in composition from the ester types, ie, epoxies, phosphate, amides, etc. The effect of any one of these on curing is usually pH-dependent. However, it is prudent to investigate each on a case by case basis.

HEALTH AND SAFETY FACTORS

Environmental Issues. The rubber industry has responded to concerns about the environment by developing new product forms for accelerators and other chemicals which improve industrial hygiene and minimize worker exposure to these materials by eliminating dust exposure and improving handling ease. Another important benefit for newer product forms is better, more consistent quality rubber products resulting from improved factory practices.

Nitrosamines. Findings that secondary amines, so common in rubber accelerators, can react with NO_x species to form the suspected human carcinogens, nitrosamines, have prompted active programs to develop alternative accelerators. Neither primary nor tertiary amines form stable nitrosamines and they are generally considered to be safe materials. The ability of each type of common rubber accelerator class to form nitrosamines has been summarized and depends on their 1°, 2°, or 3° nature (30).

The NO_x nitrosating agents present in the atmosphere are often due to air pollution. High surface area fillers such as carbon black absorb NO_x and liberate it during the vulcanization process. Of course, this is the process where NO_x is most likely to be in contact with the various accelerators.

Consequently, unacceptably high levels of nitrosamines have been detected near curing presses, near continuous curing lines, and in tire storage warehouses. In Germany, regulation TRGS 522 calls for a maximum concentration of atmospheric nitrosamines of 1 µg m³ air in general areas and 2.5 µg m³ maximum in problem areas.

Nitrosamines from vulcanizing agents can be controlled or eliminated by using only primary amine-based accelerators or accelerators which form stable fragments during cure. This is an active area of research and some potential new accelerators include dicaprolactamdisulfide, zinc-*n*-butyldithiophosphate, copper-*i*-propyldithiophosphate, zinc dibenzylthiocarbate (ZDBeC), and zinc dicyanatodiammin (30). None of these materials is a drop-in replacement for existing accelerators, and each requires redevelopment of the cure system to be successfully used.

Another way to minimize nitrosamine exposure is to use scavengers, which are reported to function by reacting rapidly with nitrosamines to render them harmless, but the relative effectiveness of each scavenger remains to be better quantified. Materials that have been investigated include the following (30):

- polyisocyanates (Desmodur TT and VK-Bayer)
- hydroxyl ammonium sulfate
- diethylhydroxylamine
- hydrazine sulfate
- polycarbodiimide (Stabaxol–Rhein-Chemie)
- propyl gallate
- α-tocopherol vitamin E (Ronotec 200–Hoffmann-LaRoche)
- ascorbic acid vitamin C
- amide amines (Rhenofit NC–Rhein-Chemie)
- calcium oxide (BF Goodrich)

Fillers for Rubber

The rubber compound usually requires an inert inorganic filler and small particle size carbon particle for reinforcement. The rubber polymers vary in inherent tensile strength from very high in the case of natural rubber to almost nonexistent for some synthetic polymers, eg, SBR. The fillers most commonly used for rubber compounds include carbon black, clay, calcium carbonate, silica, talc (qv), and several other inorganic fillers.

Carbon Black. This is the principal reinforcing filler used in rubber. Carbon black is made by three processes: the furnace process, the thermal process, and the channel process. Over 97% of black is made by the furnace process (see CARBON, CARBON BLACK).

The furnace process involves injecting low end fraction of crude oil, eg, Bunker Fuel C, into a heated chamber. The temperature, shape of the injectors of the oil, rate of injection, and other factors are controlled to produce black fillers of different particle size and structure. The particle size and structure control the reinforcing character of the carbon black. There are ~ 30 common grades of carbon black used in the rubber industry. There are numerous specialty grades produced, and several hundred are used in plastic, conductive applications, and other uses.

Table 9 shows the classification system for blacks most commonly used in rubber. The ASTM numbering system is based on the fundamental particle size of the black. Particle size is determined by several methods, including iodine absorption, nitrogen absorption, and light scattering.

Table 9. Classification of Carbon Blacks

ASTM number	Average particle size, nm	Old classification system
100–199	11–19	SAF
200–299	20–25	ISAF
300–399	26–30	HAF, EPC
400–499	31–39	FF
500–599	40–48	FEF
600–699	49–60	GPF
700–799	61–100	SRF
800–899	101–200	FT
900–999	201–500	MT

In addition to the fundamental property of particle size (and surface area), carbon black possesses a secondary characteristic of structure, best described as the tendency of individual particles to agglomerate or associate with one another. These two properties or characteristics of the carbon control the degree and nature of the reinforcing character of the black in rubber. The structure of the carbon black is determined by dibutyl phthalate absorption and surface area is estimated by N₂ absorption (Table 10).

Table 10. Properties of Common Carbon Blacks

ASTM number	Iodine number	DBP	CTAB	N ₂ absorption
N110	145	113	126	143
N121	121	113	128	145
N220	121	100	111	119
N231	121	92	108	117
N234	120	125	119	126
N299	108	124	104	108
N326	82	72	83	84
N330	82	102	82	82
N339	90	120	93	96
N347	90	124	87	90
N351	68	120	73	73
N550	43	121	42	42
N660	36	90	36	35
N762	27	65	29	28
N774	29	64	33	32
N990		42	9	6

Carbon blacks are arbitrarily classified into reinforcing, semireinforcing, and extender blacks. The furnace blacks from N100 through N400 are

considered reinforcing blacks based on the strength enhancement observed when they are incorporated into rubber. They are used in treads, black sidewalls, carcasses, and breaker compounds of radial passenger and truck tires. In addition they are commonly used in the covers of belts, a few select hose compounds, and other rubber goods requiring high tensile strength, tear, modulus, and good abrasion resistance. Effects of carbon blacks on the properties of NR and SBR are shown in Tables 11 and 12.

Table 11. Effects of Nonblack Fillers in Natural Rubber^{a, b}

Filler loading	Volume , parts	Mooney viscosity	Optimum cure (at 141°C), min	Modulus (at 300° o), MPa ^c	Tensile strength, MPa ^c	Elongation, ° o	Hardness, Shore A	NBS abrasion (ASTM D1630)	Rebound, ° o
pure gum		21	15	2.1	24.2	680	43		81.2
barytes	30	27	20	2.1	15.0	640	49	22.1	80.5
	50	24	20	2.1	11.5	610	56	19.4	
	30	32	20	2.1	15.0	640	50	24.1	81.5
ground whiting	50	43	15	1.9	10.7	590	57	21.6	
soft clay	30	35	25	7.2	17.9	500	49	44.4	81.5
	50	41	25	12.4	16.4	390	58	36.4	
hard clay	30	29	25	9.2	19.9	500	55	54.4	68.2
	50	31	25	14.9	18.4	370	64	43.8	
Purecal U	30	37	15	5.0	19.0	580	53	42.2	75.9
	50	46	15	6.1	14.9	540	57	31.8	
Calcene TM	30	32	15	3.7	19.2	640	48	50.0	70.5
	50	42	15	5.1	16.6	610	58	39.6	
zinc oxide	30	22	20	5.2	21.8	600	55	54.1	73.1
	50	23	20	5.5	15.4	550	64	48.1	
Zeolex	20	23	10	3.5	22.6	660	49	51.4	
	30	65	10	5.3	19.3	600	53	46.0	67.3
	40	62	10	7.9	18.0	540	64	39.4	
Silene EF ^d	20	47	10	4.0	24.8	670	52	59.7	
	30	51	10	5.2	20.1	610	61	48.9	67.3
	40	56	10	7.9	17.4	540	67	45.6	
HiSil 233 ^e	20	64	10	5.6	27.2	640	58	57.2	77.5
	30	81	10	9.2	25.3	590	73	60.9	65.9
	40	89	10	13.0	23.9	500	78	79.6	53.9

^a Ref. 31.

^b Recipe, in parts by wt: smoked sheets, 100.00; zinc oxide, 5.00; filler, as indicated; nondiscoloring antioxidant, 1.00; MBTS, 1.00; TMTD, 0.10; sulfur, 2.75; stearic acid, 3.00.

^c To convert MPa to psi, multiply by 145.

^d Silene EF stocks contain 2.4, 3.8, and 5.0 parts of diethylene glycol, respectively.

^e In addition to the 1 part of MBTS shown in recipe, the 20 vol HiSil 233 stock contains 0.25 parts TMTD and 0.5 parts triethanolamine; the 30 vol stock contains 0.5 parts TMTD and 2 parts triethanolamine; and the 40 vol stock contains 0.5 parts TMTD and 3 parts triethanolamine.

Table 12. Effects of Nonblack Fillers in SBR^{a, b}

Filler loading	Volume , parts	Mooney viscosity	Optimum cure (at 141°C), min	Modulus (at 300° o), MPa ^c	Tensile strength, MPa ^c	Elongation, ° o	Hardness, Shore A	NBS abrasion, ASTM D1630
pure gum		26	45	0.6	1.4	680	30	13.3
barytes	30	37	45	1.6	3.8	620	46	15.2
	50	48	45	2.3	3.8	530	54	15.3
	30	40	45	1.2	4.0	620	48	16.5
ground whiting	50	47	45	1.4	5.0	680	50	15.7
soft clay	30	42	90	2.1	9.0	1080	42	28.8
	50	51	90	2.8	7.0	1130	46	33.5
hard clay	30	41	90	2.8	11.4	890	50	34.2
	50	42	90	3.5	10.1	960	51	37.2
Calcene TM	30	45	45	1.6	12.6	860	45	26.6
	50	61	45	1.9	11.9	830	49	16.0
Purecal U	30	46	45	1.7	14.0	710	45	27.7
	50	65	45	2.5	15.2	680	52	27.3
zinc oxide	30	28	45	1.4	14.5	840	45	30.7
	50	39	45	2.1	14.9	760	54	25.6
Zeolex	20	50	30	2.1	14.0	680	47	34.1
	30	62	30	3.3	16.2	670	53	33.7
	40	79	30	4.2	14.0	620	56	32.1
Silene EF ^d	20	58	15	5.4	14.1	520	56	34.2
	30	72	15	5.3	15.9	570	60	42.0
	40	77	15	6.6	15.0	510	63	46.7
HiSil 233 ^e	20	53	60	4.0	18.7	610	51	48.7
	30	81	30	5.8	23.5	610	56	45.9
	40	112	30	7.7	21.9	560	67	40.3

^a Ref. 31.

^b Recipe, in parts by wt: SBR 1502, 100.0; zinc oxide, 5.00; filler, as indicated; nondiscoloring antioxidant, 1.00; MBTS, 1.50; TMTD, 0.10; sulfur, 3.00; Cumar MH 2.5, 15.00; stearic acid, 1.00.

^c To convert MPa to psi, multiply by 145.
^d Silene EF stocks contain 2.4, 3.8, and 5.0 parts of diethylene glycol, respectively.
^e The HiSil 233 stocks contain 2.3, 3.5, and 4.8 parts of diethylene glycol, respectively. The 30 vol stock contains 1.2 parts MBTS and 0.15 parts TMTD instead of the combination in the recipe; the 20 and 40 vol stocks contain 1.5 parts MBTS and 0.1 parts TMTD.

The N500 through N700 series blacks are considered semireinforcing blacks. These are used widely in bias tire carcasses, beads, fillers, squeegee, and other tire components not requiring the ultimate in strength and abrasion. The N800 and N900 series thermal blacks are viewed as extender blacks due to the limited reinforcement they provide for the rubber compound. In addition to the reinforcing and semireinforcing black classification, carbon blacks are often referred to as hard and soft blacks. The hard blacks are the N100 through N300 series blacks; the rest are considered soft blacks.

Silica. The main uses of silica are in the treads of off-the-road tires for improved chunking and tear resistance and as a component of the bonding system for brass and zinc-plated steel cord. These are commonly used in radial passenger and truck tire belt skim stock. In addition the body plies of steel radial truck tires, hoses and belts, and footwear use significant volumes of silica as a reinforcing filler.

Silica is an expensive filler compared to carbon black and clay. With the entrance of four or five new suppliers the capacity is more than adequate for the foreseeable future. Much development work is being done by Degussa and PPG to develop products equivalent to the Rhône Poulenc HDS silica. Huber reportedly has the ability to produce the high structure already.

Clay. Clays are a principal filler for many rubber goods. The natural air-floated and water washed clays are considered semireinforcing fillers for most rubber compounds. They are used in large volumes by the footwear, mat, low speed tire, and mechanical goods manufacturers. In addition a small amount of clay is used by the tire industry in innerliners and white sidewalls, using the clay as an extender filler in low to medium dynamic stress applications. Clays are also used as partitioning agents in slab dips and antitacks. Clays were formerly used as fillers for bias tire carcasses, however radial tires require better flex fatigue resistance than clay can impart.

Silane-modified clays have found broad acceptance in a wide variety of rubber goods. The silane serves as a bridge between the organic rubber molecule and the inorganic filler particle. The result is a filler with higher reinforcing properties than regular clay. However, the dynamic properties and the degree of reinforcing improvement usually does not permit their use as a carbon black substitute or replacement.

There are numerous companies that mine, modify, and supply clays to the rubber industry. Huber and ECC are considered leaders in the field; RT Vanderbilt is also a supplier. There are numerous other national and local sources (see CLAYS).

Processing Agents

To improve processing and to plasticize the rubber compound, numerous processing agents have been used over the years, eg, petroleum and ester plasticizers, resins and tars, liquid rubber peptizers, peptizers, fatty acids and derivatives from vegetable oils, and polyethylene and hydrocarbon waxes.

Petroleum plasticizers are the most universally used plasticizers for all rubber compounds. They improve flow and processing characteristics and also reduce the cost of the final compound. Aromatic process oils are the most common plasticizer for general-purpose rubbers. Naphthenic process oils are used where a minimum of discoloring or staining is required. Aliphatic process oils are used in the low polarity rubbers where high loadings and good solubility are required to avoid bleeding and maximize the plasticizing action.

Ester plasticizers are used mainly in very polar elastomers, such as neoprene and nitrile rubber, to improve low or high temperature performance or impart particular oil or solvent resistance to a compound; 5–40 parts are commonly used (see PLASTICIZERS). Resins and tars are added to impart tack, soften the compound, improve flow, and in some cases improve filler wetting out, as is the case with organic resins in mineral-filled SBR. Resinous substances are also used as processing agents for homogenizing elastomer blends.

Stearic acid is used in many compounds as a cure activator and in addition the stearic acid acts as a processing agent to improve mill and processing equipment release properties. Thus stearic acid may be the most widely used processing agent in the rubber industry. Stearic acid is usually used at 2–3 phr, often as a salt, and due to its low molecular weight, has a limited solubility in many elastomers (Table 13). The term stearate often refers to a fatty acid mixture containing palmitates and small quantities of other fatty acids including some unsaturated fatty acids in addition to the major component stearates.

Table 13. Fatty Acids in Rubber Compounding

Fatty acid	Mol wt	Solubility	Concentration, phr	Use
zinc unsaturated fatty acid salt	650	moderate	2–3	physical peptizer process agent
calcium stearate	650	low	1–3	release agent
sodium stearate	325	very low	2–5	external release agent
fatty acid ester waxes	500–700	low–moderate	1–5	lubricant, anti-degradent release
polyhydric ester	1500	moderate–high	2–5	process and release agent, dispersing aid

Processing agents can be prepared by blending fatty acids, zinc, and calcium soaps into waxes, such as myricyl palmitate, to overcome solubility limitations of individual products. The result is a blend of organic fatty acids, usually in paste form which serves to improve processing. In some cases, processing agents contain blends of fatty acid chemicals and include an ester-type plasticizer. This makes the processing agent more effective in highly polar elastomers such as nitrile rubber (Table 14).

Table 14. Processing Agent Properties

Material	Manufacturer	Sp gr	Mp, °C (form)	Composition
CS 68	Crystal Soap	1.10	(paste)	blend of acid soap
D 148 dry	Ventron	0.95	70 (powder)	zinc soaps of fatty acids, 75% calcium silicate carrier
Dynamar PPA 790	3M Inc.	2.0	65–90 (powder)	fluorocarbon wax blends
Dynamar PPA 791	3M Inc.	1.6	60–85 (powder)	fluorocarbon wax blends
IL 1	R. T. Vanderbilt	1.16	(powder)	sodium alkyl sulfate
IL 2	R. T. Vanderbilt	1.00	(powder)	blend of fatty alcohols
Poly Pross light dark	Disco Inc.	1.20 ± 0.02	60 (chips)	fatty acid blends
Rubars	DuBois Chemicals	1.04	60–85 (bars)	saponified fatty acids
Struktol WB 16	Struktol	1.00	95 (powder)	calcium salt of saturated fatty acid
Struktol W 33	Struktol	1.20	50–60 (flakes)	blend of fatty alcohols, acid esters, acid soaps

Antidegradants

Good aging properties of rubber compounds are essential for providing acceptable service life of rubber products. The type of elastomer used is the principal factor in considering aging properties. In general, the more saturated the main chain of elastomer, the better are the aging properties. For example, EPDM and butyl rubbers are quite stable as compared to more unsaturated polymers such as polyisoprene, SBR, and polybutadiene. Unsaturated sites are susceptible to oxidation. Antioxidants and antiozonants extend the service life of products. This discussion focuses on the use of antidegradants in general-purpose diene rubbers using conventional sulfur accelerator systems.

Factors Affecting Aging of Rubber.

Oxygen. Addition of only 1–2% combined oxygen is sufficient to cause significant deterioration of properties for most general-purpose elastomers. Oxidation proceeds by free-radical mechanisms and leads to chain scission and cross-linking (32). In chain scission, free radicals attack the polymer backbone causing softening and weakening. This is the primary effect observed for natural rubber and butyl rubber oxidation. Brittle stocks result from attack at cross-links, resulting in formation of new cross-links. This is predominant with SBR, neoprene, nitrile, and EPDM.

Loss of elongation is the most sensitive criterion for aging measurement regardless of mechanism, and it is favored over measurement of tensile loss for cured compounds. In synthetic rubber production (SBR, in particular), viscosity increases with aging and can affect processing if not prevented.

Heat. As expected, heat accelerates oxidation (33). Therefore, the effects described previously are observed sooner and are more severe as temperature is increased. Because oxidation is a chemical reaction, an increase of 10°C in temperature almost doubles the rate of oxidation.

Flexing. Flex cracking is a common mode of fatigue failure in rubber compounds. Flex cracking involves not only a mechanical fatigue but also an oxidation that is accelerated by heat generated during flexing. The higher the rate of flexing, the higher the heat generation, and the worse is the fatigue life.

Ozone. Although ozone concentration in the atmosphere is only a few ppm, it reacts with rubber double bonds rapidly and causes cracking of rubber products, especially under stress (stretching and bending). For every rubber compound, there is a critical elongation below which cracks do not form. Usually, this is in the range of 5–10% elongation.

Ozone (qv) reacts with double bonds so rapidly that it has no chance to diffuse into the rubber and therefore all action is at the surface. Thus surface-protective agents are most useful against ozone attack. For example, waxes that bloom to the surface of rubber to form an inert film are used effectively for static protection (34).

Antiozonants (qv) also protect rubber surfaces by the formation of a protection layer on the surface of rubber by reaction of the antiozonant with ozone, the ozonides (4) formed in reaction with the rubber. Certain polymers also provide good ozone protection. The use of 20–30 parts of EPDM, a low diene rubber, in natural rubber or SBR compounds significantly increases ozone resistance.

Light. Ultraviolet (uv) light promotes free-radical oxidation at the rubber surface which produces discoloration and a brittle film of oxidized rubber (35). This skin cracks in random directions to form a pattern called crazing, which can be prevented by the addition of carbon black fillers or uv stabilizers. Black stocks are more resistant to uv light than are gum or light-colored stocks. Nonblack compounds require larger quantities of nonstaining antioxidants which should bloom to the surface as the surface uv stabilizers deplete.

Sulfur. Low sulfur stocks and EV sulfur-accelerated systems have better aging resistance. Normally, the oxidation rate increases with the amount of sulfur used in the cure. The increased rate may be due to activation of adjacent C–H groups by high levels of combined sulfur. Saturated sulfides are more inert to oxidation than allylic sulfides. Polysulfidic cross-links impart excessive hardening of SBR as compared to more stable monosulfidic cross-links.

Metals. Transition-metal ions, such as iron, copper, manganese, and cobalt, when present even in small amounts, catalyze rubber oxidative reactions by affecting the breakdown of peroxides in such a way as to accelerate further attack by oxygen (36). Natural rubber vulcanizates are especially affected. Therefore, these metals and their salts, such as oleates and stearates, soluble in rubber should be avoided.

Although some antioxidants (qv) are active against catalyzed oxidation of rubber, in general the standard antioxidants do not give protection against the metal ions. Because the activity of the metal depends on its being in an ionic form, it is possible to protect stocks by incorporating substances which react with ionic metals to give stable complexes.

Selection of Proper Antidegradant. Because the various antioxidants function by different mechanisms, an antioxidant under one condition may become an oxidation promoter in a different condition. Therefore, an antioxidant should be carefully selected depending on service requirements. Most antioxidants are either amines, phenols, or phosphates. The following are some important properties in the selection of proper antidegradant that should be considered.

Persistence in Service (Volatility, Leachability, etc). Volatility is related to molecular weight and type of antioxidant molecule. In general, the higher the molecular weight the less the volatility (37). The type of molecule also has an effect. For example, hindered phenols have higher volatility than amine types of the same molecular weight. Volatility is also important from the standpoint of loss of antioxidant during service and depends on surface exposed, temperature of service, etc. Comparative loss studies for amine and phenolic-type antioxidants have been done (38). Rubber testing is affected by the volatility of antidegradants and widely varying results are obtained depending if open or closed aging chambers are used.

Solubility. Another desirable property of a degradant is its high solubility in rubber but poor solubility in water and solvents that come in contact with rubber. Poor solubility in the rubber means that only small quantities of antioxidants can be dissolved without producing a bloom. As an example, N,N'-diphenyl-*p*-phenylenediamine (DPPD) has limited use because of its poor solubility in rubber. On the other hand, phenolic and phosphite antioxidants have high solubility and bloom is not a problem.

Solubility of antiozonants in rubber is especially important because their effectiveness depends on the use of 2–5 phr which must be soluble in rubber and must migrate to the surface.

Protection Against Flex Cracking. Most antioxidants including waxes provide oxidation protection under static conditions. However, most of them are not effective in rubber products subjected to dynamic flexing, eg, sidewall compounds in tires. The best dynamic protection is provided by either N-alkyl-N'-phenyl or diaryl-*p*-phenylene diamines.

Stability. In order to have maximum effectiveness over long periods of time, an antioxidant should be stable upon exposure to heat, light, oxygen, water, etc. Many antioxidants, especially in the presence of an impurity when exposed to light and oxygen, are subject to oxidation reactions with the development of colored species. Alkylated diphenylamines are least susceptible and the *p*-phenylenediamine derivatives the most susceptible to direct oxidation.

Physical Form. For compounders, physical form is an important characteristic. They prefer solid, free-flowing, nondusty materials whereas polymer manufacturers prefer materials that are liquid and easily emulsified. Undesirable are semicrystalline materials which may stratify during storage. Also, substances to be avoided are highly viscous liquids and low melting resins which block upon storage.

Staining Properties. For white or light-colored rubber products, nonstaining antidegradants are required. For rubbers containing carbon black, discoloration is not a problem and the more potent staining amine antioxidants are used. In general, phenolic antioxidants are nondiscoloring and are preferred for gaskets and weatherstripping where migration staining is a problem.

Cost. Price is an important consideration for most commodity types of rubber products, eg, tires, hoses, and belts. For specialty items, eg, rubber products for medical applications, cost is not a primary factor; performance is more important.

Metal Deactivation. Compounds capable of forming coordination complexes with metal ions are needed for this purpose. A chelating agent such as ethylene-diaminetetraacetic acid (EDTA) is a good example.

Condition of Use. The selection of suitable antioxidants is highly dependent on the conditions under which the rubber product is to be used. It is difficult to find a single material that meets all requirements of a product. Consequently, quite often blends of antioxidants are used. Table 15 provides a summary of characteristics of commercially important antidegradants (39).

Table 15. Antidegradants^a

Type	Example (trade name)	Staining	Resistance to				
			Oxyge n	Heat	Flexing	Metal catalysis	Ozon e
hindered phenols	2,6-di- <i>t</i> -butyl- <i>p</i> -cresol (CAO-1)	none to slight	F	F	F–P	P	P
hindered bisphenols	2,2'-methylenebis-(4-methyl-6- <i>t</i> -butylphe nol) (Antioxidant 2246)	none to slight	G–F	G–F	F–P	G	P
hindered thiobisphenols	4,4'-thiobis(6- <i>tert</i> -butyl- <i>m</i> -cresol) (Santowhite Crystals)	slight	G–F	G–F	F–P	F	P
hydroquinones	2,5-di(<i>tert</i> -amyl)hydroquinone (Santovar A)	none to slight	G–F	F	F–P	P	P
phosphites	tri(mixed mono- and dinonylphenyl)-phosphite (Polygard)	none to slight	G–F	F	F–P	P	P
diphenylamines	octylated diphenylamine (Cyanox 8)	slight to moderate	G–F	G–F	F	P	P
naphthylamines	phenyl- α -naphthylamine (Akrochem PANA)	moderate	G	G	G–F	F	P
quinolines	polymerized 2,2,4-trimethyl-1,2-dihydroquinoline (AgeRite Resin D)	slight to moderate	G	E	F	P	F–P
carbonylamines condensation products	reaction product of diphenyl-amine acetone (BLE-25)	considerable	G	E	E–G	P	F–P
<i>para</i> -phenylene-dia mines	mixed diaryl- <i>p</i> -phenylenediamines (Wingstay 100)	considerable to severe	E–G	E– G	E–G	E	E–G

^a E = excellent, G good, F fair, and P poor.

Antioxidant Types. Commercially available antioxidants may be divided into three general classes: secondary amines, phenolics, and phosphites.

In general, the amines are more active than the phenolics which are in turn more active than the phosphites. Amine antioxidants, however, often cause staining problems and are therefore used mainly in black stocks. The phenolics and phosphites are relatively nonstaining and are normally used in light-colored rubbers.

Many antioxidants in these classes are volatile to some extent at elevated temperatures and almost all antioxidants are readily extracted from their vulcanizates by the proper solvent. These disadvantages have become more pronounced as performance requirements for rubber products have been increased. Higher operating temperatures and the need for improved oxidation resistance under conditions of repeated extraction have accelerated the search for new techniques for polymer stabilization. Carpet backing, seals, gaskets, and hose are some examples where high temperatures and or solvent extraction can combine to deplete a rubber product of its antioxidant and thus lead to its oxidative deterioration faster (38,40).

Among the newer techniques for providing increased protection against oxidation under severe temperature and solvent extraction conditions, is the concept of the polymer-bound antioxidant. Chemical attachment of the antioxidant molecule to the polymer backbone eliminates volatility and extraction by making the antioxidant essentially infinite in molecular weight after curing. Two general methods of accomplishing this attachment have been used: (1) functionalized macromolecules react with antioxidant molecules to provide the bound antioxidant, and (2) antioxidant molecules containing polymerizable functional groups may be copolymerized with conventional monomers during polymer production. A very good account of polymer-bound antioxidants has been reviewed (41).

Test Methods. The most common methods used for determining the efficiency of antioxidants in vulcanized elastomers involve accelerated aging tests rather than oxygen absorption experiments. The standard tests involve aging "dumbbell specimens" cut from tensile test sheets under different conditions of time, temperature, and oxygen content. They are as follows: (1) air oven (ASTM D573) at 70°C for 1, 3, 5, 7, and 14 days, or in a test tube (ASTM D865) to eliminate cross-contamination; (2) oxygen pressure (oxygen bomb, ASTM D572), at 70°C under oxygen pressure of 2 MPa (300 psi) for 2, 48, 72, and 96 hours; and (3) air pressure heat test (air bomb, ASTM D454) at 126.7°C under 0.55 MPa (80 psi) of air for 3, 5, 8, 12, 20, and 30 hours.

Because of the simplicity of the first test method, most of the comparisons are made using this technique. The effects of the aging process are usually measured on tensile properties such as tensile strength, elongation, and stress (modulus) at 300° elongation (42).

In general, one day of oven aging at 70°C corresponds to one year of natural or shelf aging (a minimum requirement for rubber products), whereas the oxygen and air bomb methods are more drastic. By varying the amounts and types or combinations of antioxidants the relative effectiveness of these materials against normal oxygen deterioration can be determined.

Tire Compounding

The pneumatic rubber tire is a torodial air container that supports a vehicle and translates the forces of driving and braking. A tire must be capable of carrying high loads, providing necessary tractive forces, withstanding high impact forces, and cushioning the vehicle and the driver. It must be highly flexible to absorb and translate many deformations per second in high speed vehicles. The forces that tires are subjected to are both complex and severe. Within a single component the tire rubber compound undergoes compression, extension, and torque forces simultaneously, or nearly simultaneously, in a rather small volume, such as a tread block segment in aggressive all-weather designs.

The pneumatic tire consists of two basic areas: the tread area which is responsible for ground contact, and the casing which is responsible for supporting load and transmitting power to the tread area. Each of these areas has several components with markedly divergent properties which serve specific and unique functions, and all of which must interact with integrity for maximum performance.

The role of the rubber compounds which are used in these basic components is threefold: (1) to provide the contact area between the vehicle and the surface; (2) to provide the cohesive material that holds the tire together such that it acts as an integral unit; and (3) to provide protection for the ultimate strengthbearing components, ie, the textiles, steel beads, and steel breakers in steel belted radial tires.

Component Parts. Desired properties in the components of a radial steel belted passenger car tire (PCT) are as follows.

The tread is designed and compounded for abrasion resistance, traction, low rolling resistance, and protection of the carcass. It often is divided into two subcomponents to maximize performance; the outer tread for surface contact, and the undertread for tying into the carcass while reducing tire rolling resistance through decreased hysteresis.

The steel belt, which provides strength and protection for the ply or plies, is encased in a compound that must possess adhesion to the steel which provides stress transfer from the very rigid steel to the many times more flexible tread, sidewall, and textile carcass components. The wedge compound is formulated to reduce belt-edge sheer stresses while tying the belt to the carcass and reducing hysteresis.

The carcass ply plies coat compound functions are basically the same as the steel breaker compound. Normally in the steel belted PCT the ply is textile cord of polyester or rayon fabrics which are soft and flexible. The truck radial steel tire normally uses a steel cord ply. Earthmover tires are of two basic constructions, ie, radial using steel and bias using textiles (see TIRE CORDS).

The sidewall compound is compounded to protect the ply and must possess resistance to weathering, ozone, abrasion, and tearing while providing excellent flex fatigue resistance. The innerliner compound must provide good air-permeation resistance and resist moist hot air aging. The inside of a hot

rolling tire resembles a steam bomb in very severe service.

The apex (often referred to as bead filler) compound must be formulated for excellent dynamic stiffness to facilitate stress distribution and provide good car handling properties. The bead insulation compound must possess good adhesion to this most important component for enclosing the plies of the tire and holding the tire to the rim. The chafer rim strip compound protects the plies from rim abrasion and seals the tire to the rim.

Tire Component Compound Properties. Table 16 summarizes desired properties of three tire components in generalized terms.

Table 16. Desirable Properties of Tire Components

Property ^a	Tread	Component	Sidewall
wear resistance	+	na	0
traction	+	na	na
rolling resistance			
cut resistance	+	na	0
adhesion (to other components)	+	+	0
cut growth resistance	+	+	+
resistance	+	na	+
hysteresis (tan δ)			
resilience (rebound)	+	+	+
heat resistance	+	+	+
flex fatigue	+	+	+
tear resistance	+	+	+
hardness	0	+	0
modulus (stretch 100% elongation)	0	+	0
modulus (compression)			
static	0	+	0
dynamic	+	+	0
processing properties	+	+	+
Mooney viscosity			
rheometer (rate state of vulcanization)	v	v	v
swell			
tack	+	+	+

^a 0 = nominal, + = maximized, - = minimized; na = not applicable, and v = varied.

There are many ways to measure these properties and some of them are proprietary. However, most laboratory tests are standardized by American Standard Testing Methods (ASTM). Many of them are interactive to various degrees. The rate and state of vulcanization is especially important to consider for components of heavier and thicker tires. The heat used to vulcanize the tire in a mold under pressure requires time to penetrate from both sides of the giant tire to the innermost portions. Securing a balanced state of cure, ie, the maximizing of physical properties in all the components, results in the innermost components having a faster rate of cure. The peripheral compounds should have a cure system which holds its physical properties well when overcured.

Most tire companies utilize their own physical and chemical laboratories in addition to extensive tire testing facilities which include laboratory and field vehicle tests. These facilities are augmented by the availability of private commercial companies which operate worldwide.

MATERIALS

Tire compounds contain the following generalized ingredients in the approximate proportions noted. Industry practice is to formulate starting with 100 parts by weight of the rubber (phr). Table 17 offers examples for various treads.

Ingredient	Parts hundred rubber (phr)
rubber (elastomers)	100
reinforcing fillers	30–90
softeners extenders	5–40
processing aids	0.1–15
protection agents	3–7
vulcanization agents	3–10

Table 17. Typical Tread Compound Formula and Ranges of Materials

Compound	Passenger car radial (PCT)			
	Standard	High performance	Low rolling resistance	Medium truck radial (MTR)
polymer	50–100 SBR	0–100 VSBR ^a	0–40 NR	60–100 NR
	0–50 BR	0–20 BR	60–100 SSBR	0–40 BR
carbon black	N343, N339, N299	N121, N134, N343, N234	N351, N343, N299	N121, N220, N110
carbon black parts	70–90	80–90	40–60	40–60
oil parts	30–60	25–45	0–30	0–20
antioxidant parts	1–2	1–2	1–2	1–2
	1–4	1–3	1–3	1–3
processing traction	0–5	0–10	0–2	0–2
sulfur	1–2	1–2	1–3	1–3
accelerator	1–2	1–2	1–2	1–2
stearic acid	1–2	1–2	1–3	2–4

^a VSBR = ESBR, SSBR, SBR, or a combination thereof.

There may be as many as 12 different compounds in a single tire. Each compound uses 8–12 ingredients. There are approximately 750 different chemicals or materials used in the tire compounding field with half the number used commonly throughout the world. Thus, the potential combinations permutations are extremely high. In addition, interactions between materials are encountered and must be controlled. For example, premixing of certain vulcanization ingredients must be avoided because of their reactions with each other and formation of end products that do not give the desired rate and state (number of elastomer cross-links) of cure in the finished tire.

Interactions are not limited to the 8–12 ingredients within a single formula, but may also occur between the compounds of the different components, through the processes of mixing, extruding, calendering (coating of steel and textiles), in pre-building compilation of components for the tire building equipment, in storage of green (uncured) tires, in vulcanization of the tire, and in the warehousing and during operation of the tire on the vehicle.

Migration of chemicals from adjoining compounds, in addition to preferential absorption of various materials into each other, makes the tire compounder's knowledge of the chemistry and physics of materials most important in realizing specific, highly varying properties in the different compounds components necessary to maximize tire performance and costs. Some interactions of materials are desired. Without migration of elastomer polymer ends across compound component interfaces, adhesion of one component to another is significantly weakened and intercomponent separation is enhanced.

Solubility parameters of materials are important in processing of compounds components. A material used in excess of this parameter can bloom (move to the surface of the component) and render it difficult in final tire assembly or block desired migration and weaken component interfaced adhesion.

Tire Elastomers. The rubbers used in various tire components are shown in Table 18.

Table 18. Rubbers Used in Various Tire Components

Elastomer	Tire types	Tire components							
		Tread	Sidewal l	Wire breaker	Ply coat	Apex	Chafer	Bead	Liner
natural rubber (NR)	all	+	+	+	+	+	+	+	+
styrene–butadiene (SBR)	most	+	+		+	+	+	+	
butadiene (BR)	most	+	+	+	+	+	+	+	
halogenated butyl (HIIR)	most		+						+
ethylene–propylene–diene monomer (EPDM)	PCT		+						

For the mechanical strength requirements of tires NR is better than SBR or BR offering excellent tread chipping chunking resistance with low tire running temperatures and improved rolling resistance. Although more difficult to mix and process it is very good in downstream processes such as extruding, calendering, and extruding, and for tire building. Although good in chip chunk-type wear and in severe stress wear, it is usually faster wearing in nominal passenger and truck over-the-road tires. It offers better ice traction and excellent flexibility in extreme low temperature service (< 40°C). It is by far the preferred rubber for wire belt and textile ply breaker coat compounds.

SBR is by far the preferred polymer for passenger car tire treads, mainly owing to its good wear and traction. It is also the least expensive polymer when produced by the emulsion (water-based) process which is the most commonly used. In the 1980s, specialty solution SBRs (SSBR) were introduced. These rubbers, using hydrocarbon solvent media and special catalytic systems, were found to offer increased capabilities for compounders to improve rolling resistance for decreased fuel consumption while maintaining high levels of wear and traction. The specialty SSBRs accomplish this through control of the styrene–butadiene polymer backbone structure, coupling of polymers, and addition of reactive chemical units at the polymer ends. Not all of these changes are used in the same specialty polymer, but coupled polymers are common and can significantly affect polymer molecular weights, ie, larger (higher molecular weights) branched polymers which are more difficult to produce, are more expensive, and have some deficiencies in processing. Many of these polymers are available in the oil black masterbatch (premixed blends) as is the emulsion SBR. SBR, whether emulsion or solution, is still not as resilient as NR or BR which limits applications in tires.

BR can be produced in both water emulsion and hydrocarbon solvent processes. It is the butadiene part of SBR that is controlled in specialty SSBRs. Originally polybutadiene found its niche in tires by offering excellent over-the-road wear characteristics in its high cis, high linearity structure. This offers resilience (low hysteresis) and excellent flex fatigue resistance along with excellent wear and could only be achieved in the more expensive solvent process. BR is more difficult to process in the factory, particularly in milling operations. It does not possess the excellent tack of NR and makes tire building more prone to problems. BR is also very low in resistance to tear (tender).

Halobutyl rubber (HIIR) is used exclusively in innerliner and white sidewalls. This rubber offers marked improvements in tire air retention owing to lower air permeability as well as excellent age resistance and flex fatigue life. The chlorine and bromine versions of isobutylene isoprene rubber (IIR) are both used to improve compatibility with the other commodity tire rubbers. It is a saturated polymer with no sites available for conventional sulfur curing. IIR offers high hysteresis at nominal tire operating temperatures and thus acts as a superior cushioning rubber. In the 1960s a tire using this rubber in the tread was commercially produced by tire producers for superior ride qualities. Its high price, high wear rate, and high fuel consumption spelled its doom in this application, but it is the preferred rubber for tire curing press bladders. A new isobutylene polymer modified with methylstyrene and subsequently bromonated is available which offers a fully saturated backbone to resist aging while improving compatibility with tire commodity rubbers. It may increase butyl presence in future tires.

Like HIIR, EPR has a fully saturated backbone and has only unsaturation points available for vulcanization cross-linking in very small percentages of the pendent diene modifier (EPDM). It is an excellent aging compound with high flex fatigue life even when heavily loaded with fillers and is utilized in PCT white sidewalls.

It is common to use more than one type of rubber within a given compound. An example of this is the truck tire sidewall which must not only possess high strength but must also have excellent aging and flex fatigue life in order to survive multiple retreading, ie, vulcanizing a new tread to the worn original tire casing. The original casing (carcass) of a radial truck tire has been known to last for a million miles while the original tread on this casing only delivers 70,000–150,000 miles.

In passenger car tires as many as four different polymers may be used for the tread compound totaling 100 phr; ie, 25 SBR 25:SSBR 30:BR 20 NR. There are at least 12 specialty SSBRs that can be purchased on the open market and some large tire companies produce their own specialty SSBRs for captive use. NR can be purchased in several grades. SBR also comes in different grades and is available in oil and oil–carbon black master batches (oil and or carbon black added to the rubber at the rubber producers) which may use a different SBR (usually higher molecular weight) than if it were produced without masterbatching.

The use of different materials within a single grouping is commonly encountered, ie, two or more different grades of carbon black, oils, etc, may be used within one compound, as well as the use of two or more rubbers making up the 100 parts total as mentioned above. Figure 14 offers an example of how properties of blended rubbers may differ from a single rubber.

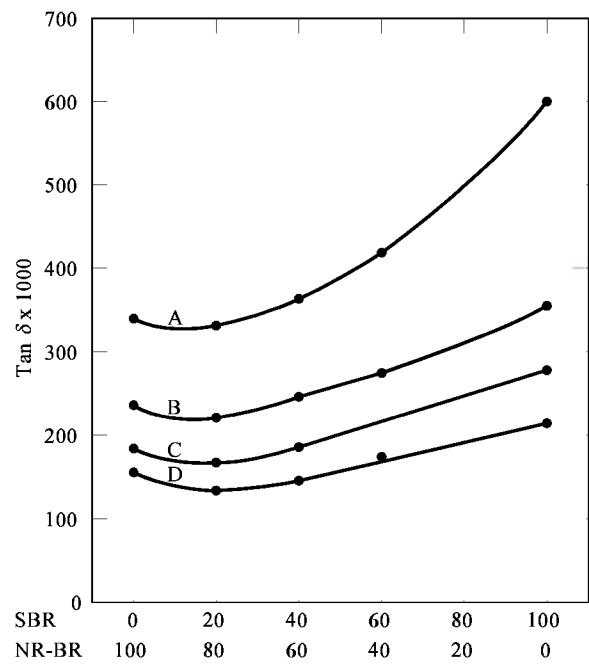


Fig. 14. Effects on compound tan δ (1 Hz, 25° deflection) triblends of SBR–BR–NR, where parts NR = parts BR and A is at 25°C; B, 0°C; C, 25°C; and D, 50°C.

Reinforcement for Tires.

Reinforcing Fillers. Carbon black is by far the most heavily used reinforcing filler for tire compounds. Annual tire usage of all grades of carbon black is estimated to be over three million metric tons annually.

Carbon Blacks. The popular carbon blacks for tires and their usage patterns are shown in Table 19.

Table 19. Carbon Blacks and Their Usage

Series	Tire types	Tire components						
		Tread	Sidewall	Ply breaker	Apex	Chafer	Bead	Liner
N100	truck	+				+		
N200	most	+				+		
N300	most	+	+	+	+	+	+	
N500	most			+			+	+
N600	most			+			+	+
N700	most			+				

In tires, carbon black is important because of the high flex fatigue and tear strength requirements of this product. Poor dispersion can result in premature cracking in both treads and sidewalls. In off-the-road tires poor dispersion results in excessive chip chunk causing a significant loss in treadwear in addition to tread lug cracking which can lead to loss of an entire lug (radial tread bar).

Because of demands for improved fuel consumption through reduced rolling resistance, a new series of carbon blacks referred to as LH, ie, N300 with this innovation would be N300 LH. Basically this series of blacks has a wider size range in both the primary particles and primary aggregates in addition to a more chemically active surface area.

Silica. Second in tonnage for tire use at an estimated 50,000 t/yr worldwide is silica, its utilization has been increasing in spite of higher per kg cost because of increased tear and lower rolling resistance. Silica as a tire reinforcing filler was first introduced through the earthmover tire line owing to its ability to improve tear resistance and lower tire running temperatures. These properties are of utmost importance in off-the-road tires which run at high operating temperatures and encounter rocks that may cut and tear out large chunks under high traction/torque conditions.

With demand for improved rolling resistance, silica usage has expanded into other tire lines. This expansion has also been boosted by the introduction of a silane coupling agent which forms a chemical bond between the silica and the rubber resulting in improved treadwear. With the use of the coupling agent silica treads are equal to some carbon blacks for wear while still offering an improvement in rolling resistance. Silane slightly reduces tear but this is acceptable in most applications. Silane can also be used with carbon black, but does not give a significant boost in wear properties. Silica has also found applications in wire breaker cost compounds where it stabilizes the copper sulfide bonding layer to the brass-coated steel cable.

Clays. Clay, which is generally considered a mild reinforcing filler, is used sparingly in tires. It is most often used in white sidewalls or in low performance tires. Clay tonnage in tires worldwide is estimated at 36,000 t annually. Clay can also be coupled to rubber with silanes, and this is the more popular version used in tires. Even with silane coupling, clays are still a weak reinforcing filler compared to both carbon black and silica.

Reinforcing Resins. Reinforcement and stiffness of a compound can also be achieved with the use of reactive resins. Resins consisting of two-component systems of resorcinol or resorcinol condensation products and a methylene donor such as hexamethoxymethylmelamine (HMMM) or hexamethyltetramine (HMT) are the most popular in tires. These materials can be prereacted and added to the formula, or for more effective results they can react *in situ*, ie, they can be added separately into the formula and react when the tire is vulcanized.

Other Reinforcement Materials. Other materials that have been used in tire compounds for reinforcement are chopped wire (brass-coated), cotton and nylon flock, chopped nylon strands, polyethylene, zinc oxide, and chopped Kevlar. Most of these materials have very limited application and some are obsolete. Others are used more extensively in solid rubber industrial tires than in pneumatics.

Softeners, Extenders, and Plasticizers. In tires the most popular softeners/extenders/plasticizers are petroleum oils. They are utilized in considerable quantities for controlling green compound viscosity and vulcanized compound hardness. The term extender is used to denote their use to allow for extended use of reinforcing fillers as the oils are significantly less expensive than are the rubbers. These materials also act as processing aids. The term plasticizer is used to denote their ability to act as an internal lubricant for processing purposes as well as for improving low temperature flexibility in the vulcanized product. The synthetic phthalates are the most effective in meeting low temperature flexibility while also serving as softeners. Cost limits their application.

The petroleum oils are of three basic types: aromatic, naphthenic, and paraffinic. Aromatic oils contain hazardous materials that require special handling precautions. Naphthenic oil does not contain hazardous levels of polynuclear aromatics (PNAs) and is less hysteretic. Because of these considerations the naphthenic oil is gaining in usage at the expense of more utilized aromatics. Paraffinic oil is only used modestly in tire compounds. The

petroleum oils are more specific to certain rubbers and therefore require compounding considerations for full optimized usage.

All three softeners migrate from one compound to another and this must be considered when compounding for a set range of stiffness (modulus) gradiations in the tire components designed to alleviate stress.

Processing Aids. Petroleum oils are processing aids as well as softeners extenders plasticizers. This section concentrates on those materials which serve mainly as processing aids without a large contribution to vulcanized properties. They are most often used at much lower levels.

Peptizers are materials used rather extensively in natural rubber (NR) compounds. When added to NR in premastication they cleave or break the molecule backbone resulting in lower molecular weight which allows for easier processing. However, there has been a move to reduce usage in order to maximize strength. Some tire companies have gone to mixing variations which are more expensive to preserve molecular weight.

Other processing aids utilized in tires are vegetable oils (fatty acids, fatty acid esters alcohols, and metal salts of these oils), naturally occurring resins such as pine tar, hydrocarbon resins from petroleum stillbottoms, and vulcanized vegetable oils (VVOs). VVOs were utilized heavily in the early 1900s, but are no longer used extensively in tires.

The main benefits of processing aids are (1) incorporation of fillers often reducing mixing energies, (2) reducing internal friction for ease of milling, extruding, and calendering, (3) compatibilizing the different rubbers for improved processing as well as for better adhesion of one compound to a different compound in the vulcanized tire (compatibilizers have gained markedly in usage because of polarity differences of blended polymers), and (4) providing improved green tack (tackifiers) for better tire building.

There is a definite trend in the supplier industry to develop multifunctional materials with value-added properties. One supplier has introduced a modified hydrocarbon resin that softens, improves filler incorporation, improves mill banding, and increases tack in the green compound while imparting increased low strain (<100%) stiffness, significantly improving tear resistance in the vulcanized compound.

Bonding Agents. These materials are generally only used in wire cable coat compounds. They are basically organic complexes of cobalt and cobalt–boron. In wire coat compounds they are used at very low levels of active cobalt to aid in the copper sulfide complex formation that is the primary adherence structure. The copper sulfide structure builds up at the brass rubber interface through copper in the brass and sulfur from the compound. The dendrites of copper sulfide formed entrap the polymer chains before the compound is vulcanized thus holding the rubber firmly to the wire.

Vulcanizing Agents. Tire compounds are almost exclusively cured (cross-linked) with sulfur. Sulfenamides, thiazoles, thiurams, guanidine, and carbamates are the most popular choices to accelerate curing. Efficient vulcanization (EV), semiefficient (semi-EV), and the conventional curing systems are all used, sometimes in the same tire for different compounds components. They are also varied depending on the rubber or rubber combinations used within a given compound, ie, a compound using a blend of NR and SBR uses a more efficient system than the NR alone may require.

Zinc oxide and stearic acid are used to activate the curing system as well as to preserve cured properties when overcuring, which is curing beyond the point of time and temperature at which maximum properties are obtained.

Overcuring is encountered to some degree in all compounds of any thickness because of the slow conductance of heat through rubber. Earthmover tires may have tire shoulder gauges in excess of 0.3 m (~12 in.). The outside inside are overcured to effect a satisfactory state-of-cure in the tread rubber.

Even though heat is being conducted to the midpoint of the tire shoulder from both sides it can take many hours to reach the same temperature as the plies and that portion of rubber next to the mold. Although the earthmover tire epitomizes this phenomena it is an excellent example of the problem of maximizing cured properties which change rather dramatically with state-of-cure. The properties of tear resistance, resiliences, modulus, tensile, elongation, and cut growth (fatigue) are significantly affected by state-of-cure. Of course many factors are involved in effecting the desired end goal of an optimized state-of-cure: the mold, the temperatures and media to effect the temperatures, the pressures, etc. To this end, a compounder's knowledge and understanding of engineering and physics, particularly thermodynamics, in addition to the chemistry of rubber and material reactions are required.

An important consideration of any vulcanization system is scorch. Scorch time is that period after temperature has been applied until a significant change in compound stiffness occurs indicating that cross-linking of the polymer has initiated. As the rubber compound is hot when the accelerators and sulfur are added, it appears that the scorch period starts immediately. As the rubber is processed through the factory to form tire components it is heated and is subject to premature cure if heat is excessive (scorch). Scorched compounds are not only more difficult to process to their finished component but some actually become impossible to process and have been known to break the steel screws of tread extruders. Scorched components also can cause premature tire failures should they end up being at the right state-of-cure for maximizing performance.

To assist in control of the onset of vulcanization, a retarder or prevulcanization inhibitor (PVI) is used. Retardation of the onset of cure does not mean that the rate of cure is slowed, in fact cure rate may actually be increased. Rather, there is an induction period prior to cure.

COMPOUND FORMULATING

Tire compounders generally use two basic approaches to formulate for achieving desired tire performance: the use of known relationships of a given compound to a specific performance parameter, and establishing the relationships of physical properties to a given performance parameter.

Using physical properties relating to performance parameters leads to the development of algorithms for predicting performance for laboratory screening of potential improvements. Many of these algorithms have been established. The two main categories of measurement criteria are quasi static and dynamic mechanical properties.

MECHANICAL STRENGTH

Strength is measured in the laboratory with tests of stretch, compression, and torque in both uncured and cured compounds.

Uncured or so-called green compounds are measured to assess processing characteristics. An example is plasticity of the green compound at a given temperature and torque to predict the ease or difficulty of milling, extruding, and calendering to secure a given profile and or gauge of a component for subsequent tire building. Cured compounds are tested for mechanical or physical properties to relate them to a specific tire parameter. Both green and cured properties are run after mixing of a formula to assure its correctness before subsequent processing for given components. Testing is also effected on the final component before tire building in some cases.

Some tests, while undergoing deformation, are usually referred to as static in that they are performed at slow speeds or low cycles. Examples of these tests are stretch modulus, ultimate tensile, and elongation to break, ie, a measure of total energy capabilities or rupture phenomena.

Dynamic properties are measured by continuous cycles of varying deformation (strain) and or stress (force required to secure a given strain), at varying frequencies which can be set close to those a component would experience in a tire. These properties are more correlative to many tire performance parameters.

The following tire performance criteria are usually associated with static mechanical strengths of the compounds: separation resistance or fuel consumption; traction in wet and dry conditions; ride, ie, cushioning and noise from harmonics; handling, ie, steering and cornering; heat separation resistance, ie, changing static properties due to heat; and wear or the changing static properties at high speeds and heat. Strain and frequency sweeps are run at varying temperatures. By the time–temperature superposition theory a 5–10°C change in temperature is equivalent to a one decade change in frequencies.

By this phenomena laboratory measurements to predict rolling resistance should be done at 60–70°C and 10–12 Hz which is closer to actual tire running temperatures for a passenger vehicle doing 88 km h (55 mi h) on an expressway, and at 0°C and 30–35 Hz for wet traction. This latter frequency is much lower than the actual frequencies encountered by a tread during braking (~10,000–20,000–Hz range), but the low temperature of testing brings it much closer to actual conditions. Increasing the tan δ or viscous modulus at 0°C results in increasing wet traction and handling.

Of course, the above applies when staying within a given tread design and at equivalent tread stiffness. Tire designs that allow water to be channeled out of the tread rubber road contact area are significantly superior to those that do not, even when more contact area is available. Tread element stiffness plays an important role in traction capabilities as softer treads have less column stiffness and reduce tread void areas, ie, there is more rubber on the road but this is offset if water channeling is significantly reduced.

For dry traction more contact is desired and the stopping distance is directly related, ie, the more contact area the shorter the stopping distance. A softer, more pliable compound conforms to the road surface topography. Too soft a compound (low mechanical strength) abrades more easily and can therefore acts as a roller and not allow sufficient contact area to be maintained. This is not readily encountered in nominal tires and conditions but has been encountered in cases of extremely high torque conditions for very fast acceleration and sudden stops.

In general, a high $\tan \delta$ at 0°C tends strongly to increase dry traction and vehicle handling through corners. However, it also directly increases the hysteretic properties, ie, the tire running temperatures, lowers rolling resistance causing greater fuel consumption and increasing the possibility of heat separation if excessive.

The elastic and viscous modulus properties of mechanical strength are used to calculate the complex modulus and the loss compliance of compounds. These two parameters are used to assess dry traction and handling. Higher values relate directly to dry traction and handling (cornering) within limits.

$$\text{complex modulus} = (\text{elastic modulus}^2 + \text{viscous modulus}^2)^{1/2}$$
$$\text{loss compliance} = \frac{\text{viscous modulus}}{\text{elastic modulus}^2}$$

Snow and wet traction are highly dependent on the tread pattern. Although the tread pattern overwhelms the compound properties in significance, the latter can play a role in optimizing snow traction. Compounds using polymers with low glass-transition temperature, T_g (-40 to -65°C), remain more flexible at low temperatures. Tread compounds with low complex modulus at 0–20°C have better snow traction.

With the advent of the wire belted radial tire, irregular tread wear became a significantly greater problem. The mechanical working of the tread elements through the rolling contact area are much more sensitive to the design and precision of the tire components as well as the wheel alignments of the vehicle. These latter mechanical entities are primary in the occurrence of irregular wear. However, the compound can help. Tread compounds with high elastic modulus at 50–75°C have been found to be less sensitive in that they increase stiffness and reduce tread element squirm or slip through the contact area. Many different conditions and correlative formulas have been derived to predict tire performance parameters from laboratory physical properties.

Aged and Fatigued Properties. Tire compounds are subjected to static and dynamic aging throughout their entire life cycle. Static aging begins even in the green (uncured) state with exposure to oxygen, ozone, light, heat, and humidity. In the tire factory, these effects can be minimized to the point where they barely exist. Factory tire compounders are assigned to tire plants to assure the integrity of their compounds throughout the production cycle. Even as tires are warehoused they are protected by regulations drafted by compounders. It is when tires are applied to the vehicle that this natural phenomena of aging becomes most important.

Heat is the primary enemy of the rubber compound. For example, the tensile strength of a tire compound at 150°C may only be 50% of its room temperature maximum. Further, most of the strength properties, such as tear and modulus, are similarly affected. When a tire is heated up by use and allowed to cool again to ambient temperature, it loses some of its original strength and is subject to attack by ozone. This can sever molecular chains, when at rest, and leads to cracking at points of deflection. Finally, flex fatigue resistance diminishes markedly upon aging and repeated cycling due to chemical changes in the compound when heated up by rolling and working. Low tire inflation means more work and more heat adding to degradation by a power factor.

Wire cords are particularly subject to degradation of their adhesion values by moisture. To combat this, halogenated butyl (HIIR) is used in tire innerliners because of its property of low air and water vapor diffusion rates. Moisture is present in most air pumps and many tires are mounted with water left in the tire on mounting. For these reasons tires and tire compounds are tested extensively at simulated aging conditions in the laboratory and on test vehicles before they are sold to the customer.

LATEX

The compounding technique for latex differs from that of dry rubber and is fundamentally simpler. A critical factor of colloidal stability makes necessary that each ingredient is of optimum particle size, pH, and concentration when added as an aqueous dispersion to the latex. Rubber latex is a colloidal aqueous emulsion of an elastomer and natural rubber latex is the milky exudation of certain trees and plants; that of greatest commercial importance is the *Hevea brasiliensis* tree.

Synthetic latices are mainly prepared by polymerizing low molecular weight monomers in a free-radical emulsion polymerization system. Artificial latices are water dispersions of reclaimed rubber, butyl rubber, *cis*-polyisoprene, *cis*-polybutadiene, or ethylene–propylene terpolymers. There are certain applications where solid rubber or a rubber solution cannot be used, eg, in preparation of foam sponge, paper coating, or saturation of a porous material. Because compounding of latex is done in the liquid state, milling or Banbury mixing is avoided. Very high application rates may be obtained, resulting in less expensive processes and in many cases continuous operations. Power requirements are usually less than for nonaqueous applications, but where thin-wall dipped rubber products undergo post-vulcanization washing, the additional cost of drying has to be taken into account. However, the use of latices avoids the hazards and additional expense associated with rubber solutions requiring toxic and inflammable solvents.

For dry-rubber compounding, the latex is coagulated, dried, and used in a solid state after mechanically inducing plasticity. Latex compounding, on the other hand, is done in the liquid state and the compounded latex used directly to form articles without being first converted to a separate solid. For such use, the latex is usually concentrated to ca 60% solids by one of several methods, the most common of which is centrifuging; a less common variant is creaming. For production of special, thin-wall dipped medical latex products, variants of the centrifuging process, such as double centrifuging or substaging, produce latex concentrate of higher purity. As with dry rubber, most uses of latex require modification by addition of vulcanizing agents and other ingredients; stabilizers are also necessary. Exceptions are adhesives in which latex is used without modification by compounding.

The general principles of latex compounding are similar to those of dry rubber. The absence of mastication in the processing of latex saves time and power, and the resulting vulcanizate can be made reasonably resistant to aging. A disadvantage of latex is the difficulty in obtaining any measure of reinforcement through use of small particle-size inorganic fillers. This failure of reinforcement is normally attributed to lack of mastication of rubber particles. Where fillers are used solely for material economy, the resulting products' physical properties are usually diminished rather than enhanced. This disadvantage is offset, however, by the fact that properly compounded and cured unfilled natural latex films can achieve tensile values in excess of 35 MPa (5000 psi) and elongation at break in excess of 950%. Some products, such as dipped rubber gloves, benefit from small additions of fillers that improve the rubber's abrasion resistance.

Another disadvantage is difficulty in drying latex rubber deposits; that, with the usually accompanying shrinkage, rules out use of latex compounds in the production of solid articles of thick section.

In order to obtain a homogenous and stable latex compound, it is necessary that insoluble additives be reduced in particle size to an optimum of ca 5 μm and dispersed or emulsified in water. Larger-size chemical particles form a nucleus for agglomeration of smaller particles and cause localized dispersion instability; particles <3 μm tend to cluster with similar effect, and over-milled zinc oxide dispersions are particularly prone to this. Water-soluble ingredients, including some accelerators, can be added directly to the latex but should be made at dilute strength and at similar pH value to that of the latex concentrate.

The particle size of typical natural rubber latex ranges from slightly higher than 1 μm to as small as 20 nm, and can be destabilized by mechanical or chemical action. Machinery used to stir or maintain latex circulation should be designed to minimize fluid shear, particularly where that may be locally intensified, eg, at the junction of blade with shaft.

Dispersions to be added to latex must have good storage stability and be compatible with the latex; the pH of each should be similar to that of the latex, eg, pH 8.5–11 for ammonia-preserved latex and pH 3.5 for cationic-preserved concentrates. Addition of low pH materials to high pH latex or vice versa generally results in mutual precipitation and coagulation of the suspended rubber particles.

Latex particles are negatively charged. Because all the charges are in the same category in varying magnitude, the charges repel particles from each other and prevent coagulation. Therefore, materials in water that produce Zn^{2+} , Mg^{2+} , or Ca^{2+} ions must be added with care to maintain latex compound stability. Insoluble dispersions should be diluted to as low a concentration as practicable at point of addition, and pH adjusted to that of the latex. Incorporation of part of a formula's stabilizer into critical additions is also of value, but must be compatible with any already used in preparation of the chemical dispersion. Controlled destabilization and deposition of the rubber by chemical action is the basis of coagulant-dipped latex product manufacture.

There are several limitations on pigmentation of latex in contrast to that of dry rubber. Materials giving polyvalent positive ions in water solution, including a number of filler materials used in dry rubber such as calcium carbonate, are avoided or used only in small amounts. Reinforcing materials, eg, the carbon black used in dry rubber compounding, have little if any reinforcing effect in latex, probably because of the much less intimate contact obtainable between the carbon and latex rubber particles in the fluid state. This can be contrasted with that obtained by action of a rubber mill on a dry rubber compound. Carbon black is, however, used in small amounts in latex as a colorant and also to ensure low electrical surface resistivity in rubber products intended for use in flammable atmospheres. Heavy pigments, such as Chrome Yellow and lead compounds, are usually avoided in latex compounding because these settle rapidly from the latex. If heavy chemicals are used the latex viscosity must be raised to minimize settling, eg, for special-purpose glove production.

Mixing of latex compounds is accomplished by stirring ingredients into the latex in the form of water solutions, dispersions, or emulsions. Although the rubber softeners needed to process dry rubber are not necessary for latex, use of emulsified softeners or polymeric plasticizers in natural or synthetic latex compounds provides lower modulus in the finished products. This reduces hand fatigue and increases touch sensitivity in dipped rubber gloves. Mineral oils are also used as an economy.

Because mill mixing is not required, and the frictional heat developed in milling dry compounds is not present, scorching is not a problem and retarders unnecessary. It is possible to design latex compounds for fast cures at low temperatures by use of ultra-accelerators and fast-curing combinations. Some (prevulcanized) latex compounds can be cured simply by allowing the dried film or article to remain at room temperature for a sufficiently long length of time. However, although there is no danger of scorch from frictional heat, such fast-curing latex compounds do exhibit change (precure or shortness, referring to loss of potential elongation at break) when kept for relatively short periods, ca 48–72 h, particularly if under agitation and at high temperature. Rubber deposited from latex in that state is grainy and exhibits poor wet strength, possibly even cracking as the gel dries during manufacture of an article.

It is possible to further increase cures by preparation of prevulcanized compounds in which, by heating the latex with the curing ingredients, the latex particles in the liquid compound develop progressive cross-linkage. After formation of a deposit or article from such a compound, it is only necessary to dry at relatively low oven temperature for a short period, or at room temperature overnight to obtain products with properties similar to those of a conventionally post-cured product. Although physical properties obtainable in such an article are not quite as good as those from conventional post-cured compound, they are adequate for many uses.

Latex dipping compounds have generally higher tensile strength than those prepared from dry rubber dissolved in mineral solvent; the latter lose some strength during mastication and milling of the rubber. Without addition of reinforcing filler, latex rubber compounds can exhibit tensile strength values in excess of 35 MPa (5000 psi). Addition of fillers can, however, improve resistance to surface abrasion, and suitable carbon black dispersions are used to reduce surface electrical resistivity of rubber.

Sulfur is the almost universal, primary vulcanizing agent for all elastomers, whether in latex or dry form. Less sulfur is normally required for latex than for dry rubber compounds. By increasing sulfur concentration to 30–50 parts by weight (30–50 parts 100 parts dry rubber content (DRC)) in latex compound, satisfactory hard rubber articles, particularly rubber–metal coatings, are obtained.

The addition of zinc oxide to the latex compound increases the rate of vulcanization and generally improves the physical and aging properties. Care is required when using zinc oxide in latex because at relatively low pH, Zn^{2+} ions form and tend to destabilize the latex compound; at higher pH, more complex ions form. It is therefore necessary to adjust the ammonia content and add stabilizers, eg, fixed alkali, casein, soaps, and surface-active agents, to prevent viscosity increase and coagulation by the zinc oxide. Zinc oxide is used as an activator in amounts of 1.0–1.50 parts (per 100 DRC) and in larger amounts as a filler, the latter imparting stiffness to many products. The maximum thickening effect of zinc oxide in latex is at 0.7 wt % zinc oxide. Thickening is minimized with low ammonia and caustic potash content. Other metallic oxides, eg, litharge and magnesium oxide, which are used in dry compounds and mineral solvent rubber solutions, are not suitable for latex compounding because of their instability. Lead compounds discolor light-colored stocks and cannot easily be kept in suspension in the latex.

Fillers, eg, clays and whiting, are used to reduce cost or provide special properties. Fillers do not reinforce rubber deposited from latex, excepting to improve abrasion resistance. They are also used to increase viscosity for latex compound spreading suitability.

The more active accelerators, particularly dithiocarbamates, are used more widely in latex than dry rubber compounding because scorching during processing is not a problem. Many articles from latex are light in color and it is desirable to keep vulcanizing temperatures below 104°C to prevent darkening. In many processes it is economically advantageous to reduce the time and temperature of vulcanization as far as possible. Dithiocarbamates are particularly suited for prevulcanization of the latex in liquid state. Zinc dibutyldithiocarbamate (ZBDC) is a much faster accelerator in latex compounds than zinc dimethyldithiocarbamate (ZMDC).

Sodium dibutyldithiocarbamate (butyl namate) (SBUD) is a water-soluble accelerator that can be added directly to the latex. Zinc dibenzoyldithiocarbamate (ZBED) is a low temperature ultra-accelerator that combines a moderate rate of cure with an exceptional resistance to precure. These properties are advantageous in dipping, in formulating adhesives, and other compounds that require frequent replenishing with fresh material. The zinc salt of 2-mercaptobenzothiazole (ZMBT) is generally used in combination with the dithiocarbamates to speed up the rate of vulcanization and impart improved aging characteristics. Tetramethylthiuram disulfide (TMTD), at sufficient level, is effective in retarding heat and sunlight deterioration of low sulfur, low modulus dipped latex films. This increases allergenic contact potential, however, if used in skin-contact products.

The use of fatty acids in latex compounds follows the same principles as in dry rubber and although natural rubber latex contains some fatty acids or their soaps, it is sometimes necessary to add more. Antioxidants are used in latex compounding as in dry rubber. Because of the light color of many dipped latex articles, nondiscoloring and nonstaining antioxidants are of particular interest. Also, because many latex articles such as gloves are thin and parts of them under some stress, they are particularly susceptible to ozone cracking. Waxes (qv), which form a protective coating by rising to the surface of the product, are used in latex compounds as with dry rubber. Other antiozonants can be used but are very discoloring and staining.

Reclaimed rubber, which is widely used in dry rubber, has little use in latex compounding. A dispersion or artificial latex of the reclaim must be made by a rather expensive process of milling in dispersing agents, eg, soaps and casein, and water. Some reclaim dispersions are used in latex compounds for such things as spread rubber goods and adhesives and fiber binders to reduce cost. However, for most latex compounds, it is not desirable because of the poor physical properties it imparts and the resultant darkening of the compound.

Synthetic. The main types of elastomeric polymers commercially available in latex form from emulsion polymerization are butadiene–styrene, butadiene–acrylonitrile, and chloroprene (neoprene). There are also a number of specialty latices that contain polymers that are basically variations of the above polymers, eg, those to which a third monomer has been added to provide a polymer that performs a specific function. The most important of these are products that contain either a basic, eg, vinylpyridine, or an acidic monomer, eg, methacrylic acid. These latices are specifically designed for tire cord solutioning, papercoating, and carpet back-sizing.

The basic constituents of all commercial emulsion polymerization recipes are monomers, emulsifiers, and polymerization initiators. Other common components are modifiers, inorganic salts and free alkali, and shortstops. The function of these different components and the mechanism of emulsion polymerization have been described (43,44).

Most synthetic latices contain 5–10 wt % of nonelastomeric components, of which more than half is an emulsifier or mixture of emulsifiers. One reason for this relatively high emulsifier concentration as compared with natural latex is that emulsifier micelles containing solubilized monomer play a principle role in the polymerization process. A high emulsifier concentration is usually necessary to achieve a sufficiently rapid rate of polymerization. Secondly, a considerable fraction of the surface of the polymer particles must be covered by adsorbed soap or equivalent stabilizer to prevent flocculation

of the latex during manufacture or subsequent use. The emulsifier levels for synthetic latex is high also because of the small particle size relative to that of natural latex.

The most commonly used emulsifiers are sodium, potassium, or ammonium salts of oleic acid, stearic acid, or rosin acids, or disproportionate rosin acids, either singly or in mixture. An alkylsulfate or alkylarenesulfonate can also be used or be present as a stabilizer. A useful stabilizer of this class is the condensation product of formaldehyde with the sodium salt of β -naphthalenesulfonic acid. All these primary emulsifiers and stabilizers are anionic and on adsorption they confer a negative charge to the polymer particles. Latices stabilized with cationic or nonionic surfactants have been developed for special applications. Despite the high concentration of emulsifiers in most synthetic latices, only a small proportion is present in the aqueous phase; nearly all of it is adsorbed on the polymer particles.

Neutral or alkaline salts, eg, KCl, K₂SO₄, K₂CO₃, or Na₃PO₄, are often present in synthetic latices in quantities of $\sim$ <1%, based on the weight of the rubber. During emulsion polymerization the salts help control viscosity of the latex and, in the case of alkaline salts, the pH of the system. Many polymerizations are carried out at high pH, requiring the use of fixed alkali, eg, KOH or NaOH. Very small amounts of ferrous salts can be employed as a component of the initiator system, in which case a sequestering agent, eg, ethyldiaminotetraacetic acid (EDTA) may be included to complex the iron. Water-soluble shortstops, eg, potassium dithiocarbamate, may also be included in very small amounts (ca 0.1 parts).

For some applications, eg, foam rubber, high solids (>60%) latices are required. In the direct process, the polymerization conditions are adjusted to favor the production of relatively large average particle-size latices by lowering the initial emulsifier and electrolyte concentration and the water level in the recipe, and by controlling the initiation step to produce fewer particles. Emulsifier and electrolyte are added in increments as the polymerization progresses to control latex stability. A latex of $\sim$ 35–40 wt% solids is obtained and concentrated by evaporation to 60–65 wt % solids.

In the second process, a small particle-size latex is prepared and treated so that a limited and controlled degree of particle agglomeration occurs. The agglomerated latex is then concentrated as before but, because of the particle-size distribution obtained, the solids may be raised to ca 70 wt %. Two methods exist for agglomeration of latices, ie, chemical and freeze agglomeration (45,46).

Styrene–Butadiene and Related Latices. The SBR latices can be classified into hot and cold types, as determined by polymerization temperatures, and subdivided into low and medium solids and high solids. A summary of commercial SBR latices is given in Table 20. Hot SBR latices are polymerized at 49–66°C and mainly supplied at medium solids content, ca 42–50 wt %. Only two high solids (ca 60–70 wt %) types, namely 2003 and 2004, are available and are used mainly for special applications such as the base for graft polymerizations for high impact plastics. Type 2006 is supplied ca 27 wt % solids for the chewing gum industry. There are a wide variety of high styrene resin latices for many diverse applications from carpet backing to paint and paper coating.

Table 20. Styrene–Butadiene Latices

Type	Monomer ratio: styrene butadiene	Mooney viscosity, ML-4	Solids, wt %	pH	Emulsifier ^a
<i>Elastomeric, hot</i>					
2000	46 54	70	39–44	10–11	RA
2001	46 54	30	39–44	11	RA
2004	0 100		58	10	RA–FA
2006	25 75	50	27	10	FA
J-9049	46 54	70	49	10	RA
J-8146	46 54	70	59	11	RA
<i>Elastomeric, cold</i>					
2105	30 70	140	62	10–11	RA–FA
2107	44 56	140	62	10–11	RA–FA
2108	25 75	>140	40	11	FA
2113	44 56	130	48	10–11	RA–FA
J-9428	25 75	135	69–72	10–11	RA–FA
5352	25 75	135	69–72	10–11	FA
<i>High styrene latices, hot</i>					
2711	59 41		53	10–11	RA
2714	82 18		54	10–11	RA

^a RA rosin acid; FA fatty acid.

There has been a marked trend toward concentration of higher styrene (ca 40%) polymers in hot latices, and lower styrene (mostly 20–30% bound styrene) types in cold latex series. This is a reflection of the fact that lowering the polymerization temperature of high styrene copolymers produces little or no gain in the physical properties of the copolymer.

Another difference between hot and cold elastomeric SBR latices is that hot types are carried to < 90% conversion and not normally shortstopped. The cold latices are usually shortstopped at ca 60–80% conversion. Again the desired physical properties of the contained copolymer are responsible for these differences. Cold latices are used in applications where the modulus, eg, in foam, or retention of physical properties at high filler loadings, eg, in fabric backing, are required. The cold latices are generally supplied at a higher solids concentration than the hot series because of these uses.

Styrene–butadiene latexes generally are quite stable mechanically because of the presence of relatively large amounts of emulsifying and stabilizing agents, and therefore require addition of less stabilizer in compounding. The applications of SBR latex are classified in Table 21. This classification indicates the scope of the industry and illustrates the large number of diverse applications in which synthetic latices are employed. The latex types previously found most suitable for particular applications are also listed.

Table 21. Applications of SBR Latices

General uses	Specific applications	Types
paper	saturation	2000, 2001, 2740
	beater addition	2000, J-9049
	coating	Dow 512R, 630, 636, 620
carpets	back sizes	2000, 2105, 2714, modified SBR
pile fabrics	coating back of pile fabrics	2000, modified SBR
fabric adhesives	fabric tie coats	2000, 2107, 2113
tires	cord dipping	2000, 2108, J-9049, Pyratex, Gen-Tac
footwear	combining fabrics	2000, 2002, 2107, 2113
asbestos fiber	brake linings, gaskets, linoleum base	2000, 2758, J-9049
jute	shoe insoles, etc	2000

jute and sisal	upholstery pads, rug underlay	2000, J-9049
foam sponge	cushioning, rug underlay	2105, J-9428, 5352
batteries	separator sheets	2000
chewing gum		2006, J-8146
paint	pigment binder	Dow 512-K, 762-W

Butadiene–Acrylonitrile Latices. Nitrile latices are copolymers of butadiene and acrylonitrile in which those copolymerized monomers are the main constituents (see ELASTOMERS, SYNTHETIC NITRILE RUBBER). The latices differ mainly in ratio of comonomer and stabilizer type. They can be classified as low and medium acrylonitrile (ACN) types. The latter contain 35–40 wt % nitrile rubber, and low types ca 27–29 wt %.

Nitrile latices are used in a wide variety of applications, including production of dipped nitrile rubber products. In the principle use of paper saturation, adhesives and fiber bonding, small particle size and optimum surface tension is desirable to achieve rapid penetration and setup or drying.

Films deposited from compounded nitrile latices can be vulcanized with sulfur and accelerators, assisted by relatively high levels (ca 4.0–5.0 parts 100 DRC) of zinc oxide. For other uses, nitrile latices are sometimes used in unvulcanized form. An application of medium solids nitrile latex, eg, Nitrex J-6849 and Polysar 845, has been in preparation of oil-resistant foams for lubricants in heavy-duty bearings, such as railroad-car journal boxes.

Neoprenes. Of the synthetic latices, a type that can be processed similarly to natural rubber latex and is adaptable to dipped product manufacture, is neoprene (polychloroprene). Neoprene latices exhibit poor initial wet gel strength, particularly in coagulant dipped work, but the end products can be made with high gum tensile strength, oil and aliphatic solvent resistance, good aging properties, and flame resistance. There are several types of neoprene latex, available at moderately high (ca 50 wt %) and medium solids content. Differences in composition between the types include the polymer's microstructure, eg, gel or sol, the type of stablizer, and the total solids content (Table 22).

Table 22. Neoprene Latices^a

Neoprene latex type	Comonomer	Emulsifiers ^b	Chlorine content, wt %	pH at 25° C	Standard solids, wt %	Distinguishing features	Primary applications
101	methacrylic acid	poly(vinyl alcohol) ^c	36	7.0	46	high strength, carboxylated polymer with outstanding mechanical and chemical stability	adhesives, bonded batts, coatings, saturants
102	methacrylic acid	poly(vinyl alcohol) ^c	36	7.0	46	medium strength, carboxy-lated polymer with outstanding mechanical and chemical stability	adhesives, bonded batts, coatings, saturants
357		potassium salts of dispropo-portionated resin acids and fatty acid and polymerized potassium salts or alkyl-naphth-alenesulfonic acid	38	12.5	61	designed specifically for foam	foam
400	2,3-dichloro-1,3-butadiene	potassium salts of dispropo-portionated resin acids	48	12.5	50	maximum chlorine content; rapid crystallizing; outstanding ozone and weather resistance	adhesives, bonded batts, coatings
450A	acrylonitrile	potassium salts of dispropo-portionated resin acids and fatty acid	31	9.5	41	maximum hot-oil and plasticizer resistance	asbestos binder, back coatings
571	sulfur	sodium salts of resin acids	37.5	12.0	50	high strength cured films combined with low permanent set	dipped goods, elastici-zed aluminous cement, coatings, cord adhesives
572	sulfur	sodium salts of resin acids	38	12.0	50	rapid crystallizing; rapid coalescence under pressure	wet laminating adhesives
601A		sodium salts of resin acids	38	12.0	60	creamed Latex 842A	mastics
635		sodium salts of dispropor-tionated resin acids	38	11.0	60	creamed Latex 735A	mastics
671		potassium salts of dispropo-portionated resin acids and fatty acid and polymerized potassium salts of alkyl-naphth-alenesulfonic acid	38	12.0	59	wet-gel strength, high solids with low viscosity, 671A has a lower viscosity than 671	dipped goods, laminating adhesives, mastics, bonded batts, impreg-nated paper, extruded thread, contact bond adhesives
671A				12.5	59		
735A		sodium salts of dispropor-tionated resin acids	38.5	11.5	45	designed for wet-end addi-tion to fibrous slurries	fiber binder
750	2,3-dichloro-1,2-butadiene	potassium salts of dispropo-portionated resin acids	40	12.5	50	high wet-gel strength, low modulus, crystallization resistance	dipped goods, cord adhesives, contact bond and
842A		sodium salts of resin acids	37.5	12.0	50	medium strength cured films having a slow crystallization rate	con-struction adhesives, dipped goods, saturants, coatings, bonded batts

^a Ref. 47.
^b Emulsifier class is anionic unless otherwise noted.
^c Emulsifier class is nonionic.

As in dry rubber compounding, neoprenes require sulfur and an accelerator for vulcanization, but zinc oxide can also be used to accelerate the cure and to function as an acid acceptor. Unlike dry neoprene compounding, no magnesium salt is used because of its destabilizing effect on the latex. Organic accelerators are used to improve the physical properties of neoprene latex films but their effect is not generally as great as when used with other polymers. The most effective accelerators in neoprene latex are thiocarbanilide (*sym*-diphenylthiourea) used either alone or in combination with DPG, or combinations of TETD and SBUD.

Use of a good antioxidant is recommended for almost all neoprene compounds; where color is not of importance, a staining antioxidant can be used. The substituted or hindered phenols, eg, Naugawhite (Uniroyal) and Antioxidant 2246 (American Cyanamid), are used where a minimum of product discoloration or staining is required.

The tendency for light-colored neoprene vulcanizates made from latex to discolor when exposed to light is markedly retarded by compounding with triglycerides of long-chain fatty acids in combination with nondiscoloring antioxidants (48). Resistance to discoloration increases with increasing amount of safflower oil up to 15 parts 100 DRC. Fillers are added to neoprene latex to modify the film's physical properties, lower compound cost, improve sunlight and water resistance, and adjust processing properties of the liquid. Fillers do not, however, improve the polymer tear strength. Hard clays or a limited amount of carbon black are used as fillers. Mineral and light process oils are used to reduce modulus of neoprene latex films, whereas ester plasticizers are used to improve low temperature properties.

Because the viscosity of neoprene latex at a given solids content is less than that of natural rubber latex, thickeners are generally needed with the former. Methylcellulose and the water-soluble salts of poly(acrylic acid) are the two most commonly used thickeners. Natural and synthetic gums are also used.

As in dry compounding, acid acceptors must be incorporated into neoprene latices because of the wide use of these latices in coating fabrics and metals. The hydrochloric acid that forms during service life has a particularly destructive effect on coated cotton fabrics that are not adequately protected. High zinc oxide concentration (ca 15 parts) and use of 0.4 parts *N*-phenyl-*N'*(*p*-toluenesulfonyl)-*p*-phenylenediamine (Aranox, Uniroyal) as an antioxidant provides adequate protection.

Vinylpyridines. The vinylpyridine latices were developed specifically for adhering rubber stocks to fibers, particularly nylon (49), but had earlier use also to improve processing and solvent resistance of dipped, lower-AN acrylonitrile latices. In general, the polymers are high diene types containing 10–15 wt % copolymerized 2-vinylpyridine and an approximately equal amount of styrene. These latices are supplied at ca 41 wt % solids and are emulsified with anionic soaps. There are three generally available vinylpyridine latices, namely Ge-Tac (General Tyre & Rubber Company), Hycar 2518 (BF Goodrich), and Polysar 781 (Polymer Corporation).

Carboxylated Latices. One advance in latex technology has been the introduction of latices in which the polymer phase contains functional groups (50). The functional groups are derived from the use of unsaturated monomers containing carboxy groups in the polymerization system. Carboxylated styrene–butadiene latices have been used increasingly in carpet-backing applications because of their self-curing feature, ie, the use of sulfur and accelerators is obviated, resulting in lower cost but more particularly very low odor compounds. Presumably carboxylated latices obtain their strength from polar carboxyl groups and fairly high styrene content. Metal oxides and melamine–formaldehyde resins can be used as curing systems. In carpet-backing applications, zinc oxide should be omitted from formulations because it can reduce tuft-retention values.

Types of Latex Compounds. For comparison with dry-rubber compounds, some examples of various latex compounds and the physical properties of their vulcanizates are given in Table 23. Recipes of natural rubber latex compounds, including one without antioxidant, and data on tensile strength and elongation of sheets made from those, both before and after accelerated aging, are also listed. The effects of curing ingredients, accelerator, and antioxidant are also listed. Table 24 also includes similar data for an SBR latex compound. A phenolic antioxidant was used in all cases.

Table 23. Properties of a Natural Rubber Compound^a

Cure rate	Property	No antioxidant	Antioxidant, 1.0 part (dry)
<i>Unaged, cured in air at 121 °C</i>			
3 min	700 ⁰ % modulus, MPa ^b	8.2	9.7
	tensile strength, MPa ^b	33.3	32.4
	elongation, %	>970	>910
15 min	700 ⁰ % modulus, MPa ^b	5.5	9.1
	tensile strength, MPa ^b	34.3	33.5
	elongation, %	>990	>920
<i>After air-oven aging 7 d at 70 °C</i>			
3 min	700 ⁰ % modulus, MPa ^b	1.9	2.1
	tensile strength, MPa ^b	35.6	37.1
	elongation, %	920	880
15 min	700 ⁰ % modulus, MPa ^b	1.6	1.3
	tensile strength, MPa ^b	32.8	34.2
	elongation, %	980	920
<i>After oxygen-bomb aging 96 h at 70 °C</i>			
3 min	700 ⁰ % modulus, MPa ^b	1.9	3.3
	tensile strength, MPa ^b	27.3	33.0
	elongation, %	900	810
15 min	700 ⁰ % modulus, MPa ^b	1.5	2.8
	tensile strength, MPa ^b	20.5	30.5
	elongation, %	940	880
<i>After oxygen-bomb aging for 7 d at 70 °C</i>			
3 min	70 ⁰ % modulus, MPa ^b	2.9	4.4
	tensile strength, MPa ^b	20.4	30.1
	elongation, %	870	820
15 min	700 ⁰ % modulus, MPa ^b	1.5	2.5
	tensile strength, MPa ^b	11.0	16.8
	elongation, %	920	840

^a Dry basis natural rubber compound recipe, in part by wt: high ammonia natural latex rubber concentrate, 100.0; potassium hydroxide, 0.5; Nacconal 90F (alkylarenesulfonate (Allied Chemical Co.)), 1.0; zinc oxide, 3.0; sulfur, 1.0; ZMBT, 1.0; zinc diethylthiocarbamate (ZEDC) (trade names: Ethazate (Uniroyal, Inc.), Ethyl Zimate (R. T. Vanderbilt), 0.3; antioxidant, as indicated. Wet-basis natural rubber compound recipe, in parts by wt: natural latex (NC 356), 167.9; potassium hydroxide, 2.5; Nacconal 90F, 5.0; zinc oxide, 5.45; sulfur, 1.65; ZMBT, 2.0; ZEDC, 2.0; antioxidant, as indicated. All films poured from freshly mixed compounds, dried overnight in place, then lifted and dried 1 h in air at 50°C before curing.

^b To convert MPa to psi, multiply by 145.

Selection of Rubbers for Specific Applications

The selection of the type of rubber to be used for a specific product depends on the technical requirements of the product, the properties attainable by compounding, and economic factors. By far the largest volume of rubber is used in the manufacture of pneumatic tires for passenger cars, trucks, aircraft, and farm machinery. The polymers used for these applications are natural rubber, SBR, polybutadiene, polyisoprene, halogenated butyl, and some EPDM rubber. Butyl, EPDM and natural rubber are also used for innertubes. The essential requirement for a typical passenger tire include resistance to abrasion, cutting, and chipping; reduced rolling resistance; resistance to cracking and crack growth; adequate flexibility at the lowest temperature encountered in service; a sufficiently high coefficient of friction between the tread and road to minimize slipping and skidding; adequate adhesion of the tread to the carcass; sufficient stability of the material with time so that excessive deterioration does not occur in the normal life of the tire; and a moderate hysteresis so that excessive temperatures do not develop in service. When compounded with 70 parts of N339 (HAF-HS) or N234 (ISAF-HS) carbon black, vulcanized with a suitable sulfenamide accelerator system, and adequately stabilized with antiozonants, the various grades of SBR, polybutadiene, and natural rubber provide these requirements. Generally, SBR and 25–35 wt % polybutadiene are used in passenger tires.

Latex Mixes. The first step in latex fabrication is to bring compounding ingredients into solution or suitable dispersion form. Most ingredients are not water soluble and it is necessary to emulsify liquids and microdisperse solids in water (see LATEX TECHNOLOGY).

Preparation of Dispersions. The general procedure for preparing dispersions is to make a coarse slurry of the powder with water that contains small amounts of dispersing agent and stabilizer, then grind the slurry in a suitable mill to the desired particle size. For most applications, 7 μm average particle size, with <2% over 10 μm, is acceptable. Care must be taken to avoid overmilling as with some materials that can result in particle agglomeration; some types of zinc oxide are particularly prone to overmilling. Water used should be as pure as possible. Dispersing agents are additives that cause the dispersed particles to repel one another, thereby keeping the particles in suspension (see DISPERSANTS). The action is attributed to increasing magnitude of the electrostatic charge on the particles. Typical dispersing agents include condensation products of sodium β-naphthalenesulfonate and formaldehyde and an alkali metal salt of sulfonated lignin. The amount of dispersing agent required is specific to the material being milled and the desired particle size, and has to be determined by experiment. The type of latex and its particle size, and the nature of the dispersing agent itself, must also be considered. It has been found that surprisingly large differences in dispersion result from minor differences in concentration of the dispersing agent (51). This study showed the effect of sodium naphthalene–formaldehyde sulfonate dispersing agent on a zinc oxide dispersion:

<i>XX-78 zinc oxide</i>							
dispersing agent, wt %	0.6	0.8	1.0	1.25	1.5	2.0	2.75
apparent viscosity, mPas(cP)	975	190	112	88	118	215	740

The particle size in many synthetic latices is much smaller than in natural rubber latex, implying a much greater surface area in relation to particle weight. This greater surface area may rob the added dispersion of some of its dispersing agent and create a degree of instability, therefore more dispersing agent should be used in dispersions added to SBR than when added to natural rubber latex. Sometimes it is also necessary to add a wetting agent with the dispersing agent to produce a satisfactory dispersion. Usually <1 wt% wetting agent is used, because an excess tends to produce foaming during the dispersing operation.

Two classes of grinding equipment are used to prepare dispersions. The first, the colloid mill, does not effect a particle size reduction but does break down aggregates of fine particles. Colloid mills are used for such powders as clays, precipitated whiting, etc. Sometimes these mills are used to process zinc oxide but for dipped rubber products that is not satisfactory.

The second class of grinding equipment is used to prepare dispersions. Typical of this class are ball and pebble mills, ultrasonic mills, and attrition mills. Solids, eg, sulfur, antioxidants, accelerators, and zinc oxide, are generally ground on this equipment (see SIZE REDUCTION). Ball mill action is assisted in some mills by a combination of dispersion circulation by an external pump and mechanical oscillation of an otherwise fixed nonrotary mill chamber. Where ball mill chambers are rotated it is necessary to experimentally establish an optimum speed of rotation, the size and weight of the ball charge, and ensure the mills do not overheat during the grinding period.

Ultrasonic mills produce dispersions by causing the slurry to impinge against a stainless steel blade at high velocity. These mills are fast and effective but have a disadvantage of being easily shut down by coarse slurry particles or large particles of foreign matter. An attrition mill functions by means of a reciprocating paddle, which violently agitates a mixture of the slurry, and a grinding medium usually consisting of pebbles. A pump also accelerates grinding by circulating the slurry. As with mechanically oscillated fixed ball mills, attrition mills rapidly produce high concentrations of fine particles.

Typical recipes for an antioxidant dispersion and an ultra-accelerator dispersion are given below. The antioxidant recipe is for Aminox (Uniroyal, Inc.), a low temperature reaction product of diphenylamine and acetone.

<i>Material, parts by wt</i>	
A. water	68.0
B. water	22.8
ammonia (28 wt % NH ₃)	1.0
Blancol (GAF Corp.)	4.0
Dowicide A (Dow Chemical Co.)	0.2
casein	2.0
C. bentonite	2.0
D. Aminox	100.0
<i>Total</i>	<i>200.0</i>
total solids, wt %	54.2
active solids, wt %	50

Procedure. Add A to ball mill. Compound B separately and add to mill. Add C and D to mill. Ball mill for 4 d, cooling with water to avoid sintering of the Aminox.

The ultra-accelerator dispersion recipe and directions are for Methazate (Uniroyal, Inc.), which is zinc dimethyldithiocarbamate.

<i>Material, parts by wt</i>	
A. water	70.0
B. ammonia (28 wt % NH ₃)	1.0
Blancol (GAF Corp.)	4.0
Dowicide A (Dow Chemical Co.)	2.0
casein	2.0
water	22.8

C. Methazate	100.0
<i>Total</i>	<i>200.0</i>
total solids, wt %	53
active solids, wt %	50

Procedure. Add A to ball mill. Compound B and add to ball mill. Add C to ball mill. Ball mill for 48 h.

Preparation of Emulsions. An emulsion is a system in which one liquid is colloidally dispersed in another (see EMULSIONS). The general method for preparing an oil-in-water emulsion is to combine the oil with a compatible fatty acid, such as an oleic, stearic, or rosin acid, and separately mix a proportionate quantity of an alkali, such as potassium hydroxide, with the water. The alkali solution should then be rapidly stirred to develop as much shear as possible while the oil phase is added. Use of a homogenizer to force the resulting emulsion through a fine orifice under pressure further reduces its oil particle size. Liquid oleic acid is a convenient fatty acid to use in emulsions, as it is readily miscible with most oils.

The addition of 2–10 wt % of castor oil (qv) or a light petroleum spindle oil to the oil phase prior to emulsification is said to improve the emulsion's mechanical stability. Slightly viscous emulsion is more stable during storage than a water-thin type. A small amount (ca 5 vol %) of less stable emulsion added to the main plasticizing emulsion (used in natural latex compounds for coagulant dipped glove production) reduces finger-web formation that can occur during mold withdrawal from the latex.

The use of fine particle size insoluble material, or diatomaceous earth-based fluid structuring agents, in the aqueous phase also stabilizes some emulsions. A recipe for an emulsion of a liquid, nondiscoloring antioxidant (Naugawhite) is given below:

<i>Material, parts by wt</i>	<i>Dry</i>	<i>Wet</i>
water (hot)		19.0
Nopco 1444B (highly sulfonated castor oil produced by Nopco Chemical Co.)	5.4	6.0
Naugawhite (75% active)	75.0	75.0
<i>Total</i>	<i>80.4</i>	<i>100.0</i>

Procedure. Add Nopco 1444B to hot water with high speed stirring. Add Naugawhite slowly, allowing a few minutes between additions. After all the Naugawhite has been stirred in, continue stirring for 15 min.

A recipe and directions for a mineral-oil emulsion is given below.

<i>Material, parts by wt</i>	
A. mineral oil	70.0
oleic acid	1.5
B. potassium hydroxide	1.5
water	27.0
<i>Total</i>	<i>100.0</i>

Procedure. Add A to B using an agitator such as the Eppenbach Homo-mixer. Put the emulsion through a homogenizer to obtain a very small particle size and a high emulsion stability.

A recipe and directions for a mineral oil emulsion is given below.

<i>Material, parts by wt</i>	
A. mineral oil (light spindle oil of good quality is suitable)	70.00
oleic acid	1.50
B. potassium hydroxide	1.50
deionized water	27.00
<i>Total</i>	<i>100.00</i>

Procedure. Add A to B using an efficient, high speed mixer. Put emulsion through homogenizer to obtain very small particle (ca 2–4 μm) particle size and a high emulsion stability.

Mixing. The mixing of latex compounds is in theory a simple operation, but it is necessary to take note of several factors to achieve a good, stable liquid latex compound. Of importance are that liquid solutions, dispersions, and emulsions should have a pH compatible with that of the raw, stabilized latex concentrate; that particle size in each dispersion should be sufficiently small and stable that these do not rob the latex of its stabilizer; and that concentration of insoluble dispersions should not be too high, as those can cause localized agglomeration at point of addition. Where mechanical stirring is used during compounding, the impeller blades should be designed to avoid excessive fluid shear, particularly at the junction with the shaft. This principle applies equally to mechanical circulation of latex compound in dipping tanks; archimedian screw impellers, set with sufficient clearance within tubular chambers at the side of the dipping tanks, is one method to move the latex compound mass with minimal shear.

Forming. For production of useful articles from latex rubber compounds, it is necessary to convert them into solids of desired form. A variety of methods are used to accomplish this. The straight-dip is the simplest of any method used in making dipped rubber articles, and is used for manufacture of very thin-walled goods from which water can rapidly be removed by evaporation. The glass, ceramic, or metal forms are dipped directly into the usually circulated latex compound and then slowly removed. To ensure uniform thickness, the former is slowly rotated after dipping while the still liquid film is drying; films can be dried at room temperature or in warm air at up to 50°C. Thicker articles can be made by a multiple-dipping process, with drying between dips; manufacture of contraceptive sheaths made up of two consecutive dips is an example. For thicker articles, multiple dipping is neither practical nor economic for most commercial rubber articles. The thickness of rubber deposited by a single straight dip can seldom exceed 0.10 mm; attempts to achieve more, by increasing latex viscosity, usually result in uneven drying and consequent cracking of the wet rubber gel.

The use of porous formers in the dipping process, or porous molds prepared from plaster of Paris or unglazed porcelain with a surface pore size smaller than the majority of rubber particles, has been widely adopted in the latex industry. With the porous porcelain formers, the rubber particles are filtered on the surface of the formers. The rubber latex coagulates because of its high concentration to form a film of increasing thickness as more water is absorbed into the ceramic. Its rate of increase diminishes sharply beyond an optimum period of time, however, depending on the various characteristics of the ceramic.

When plaster of Paris is used, the calcium ions in the plaster cause coagulation. With porous molds, the latex compound is poured through a funnel-shaped opening until the mold is full and allowed to remain in the mold until the desired thickness, determined by experimentation, has been deposited on the mold wall. The mold is then emptied of excess compound and placed in an oven at ~60° C to dry for one hour. The interior surface of the cast product is dusted with talc (qv) to prevent sticking, and after its removal from the warm plaster mold can be returned to the 60°C oven for further drying; if necessary, the product can be supported on a frame during drying. With some articles, to prevent potential interior weakness from pour lines, the molds are rotated in several planes to give the latex opportunity to uniformly distribute over the interior before setting, eg, for cast rubber dolls and toys.

Accumulation of very small stable rubber particles and previously soluble natural proteins over time reduce the absorbance of water into the mold.

Those made of plaster of Paris have a finite life for that reason, but residues can be removed from ceramic molds, by use of an inorganic acid to clean the surface, and a chelating agent such as EDTA for deeper cleansing.

Electrodeposition. This earlier method of production has been dropped in favor of simpler methods (52). Natural rubber particles in latex are negatively charged; therefore, when an electrical current is passed through the latex, the particles move toward the anode. On arrival at the anode, the particles are neutralized and deposited as coagulated rubber. Nearly all common vulcanizing agents when properly dispersed are also negatively charged and likewise deposited with the rubber on the anode. The main drawback of this method is evolution of oxygen at the anode, producing oxidation and porosity in the product.

Coagulant Deposition. The same results as those obtained by electrodeposition can be achieved more simply and economically by use of chemical coagulants, eg, solutions of polyvalent metal salts, in place of electrical current. Two different forms of this general method have been patented (53,54). In the American anode process, the impermeable former is first dipped into a coagulant solution consisting usually of calcium chloride, or a mixture of calcium chloride and calcium nitrate, in a solvent, preferably alcohol. The former is withdrawn and the solvent evaporates, leaving a thin viscous film which is not disturbed when the former is subsequently dipped into the latex compound. The thickness of the rubber film deposited depends on the dwell time in the latex and the stability and viscosity of the latex compound. Concentration of the coagulant solution has bearing on the rate of deposition, which diminishes throughout the period of immersion. Attempting to achieve too thick a deposit by extending time of immersion can result in poor, uneven drying of the wet rubber gel. This can cause formation of steam between the gel and former, distorting the product during subsequent vulcanization, particularly if the product has been formed by consecutive depositions.

When the film has attained the desired thickness and is partially dried, it is passed through hot water (ca 60–80°C) to remove coagulant and other water-soluble ingredients. The film is then dried in air at a temperature not higher than 90°C before vulcanization in the usual manner. The time in water prior to that, however, depends on the overall process. Continuous dipping lines producing rubber gloves use as little as seven minutes in water at 80°C, where the gloves are scheduled to also undergo halogenization with chlorine solution or given extended time in water after removal from the molds.

The Uniroyal process differs from that of American anode, principally in that the first dip is in the latex compound rather than in the coagulant. The resulting thin rubber film acts as a carrier for a coagulant subsequently absorbed by it. Volatile acids, eg, formic, acetic, or lactic acid, or cyclohexylamine dissolved in alcohol or acetone or both, have generally been used in this process, but in the 1990s water is more commonly used than ethanol.

For natural or neoprene latex dipping, suitable coagulants are as follow:

<i>Parts by wt</i>	<i>A</i>	<i>B</i>	<i>C</i>
calcium nitrate crystals	25	25	25
denatured ethyl alcohol	75	74.75	70.75
nonionic surfactant		0.25	0.25
release agent (calcium carbonate, etc)			4.00
<i>Total</i>	<i>100</i>	<i>100</i>	<i>100</i>

It is often more convenient to use calcium nitrate solution, commercially available at 50 wt % of an adequate purity, in making up coagulants. Use of this also enables more rapid adjustment of strength during production. A convenient, rapid measurement of coagulant strength is by Hydrometer. Dependent on level and type of release agent, use of nonionic surfactants as wetting agents may result in excessive surface foam; levels as low as 0.10 are often sufficient. Release agents and level used should be chosen with regard to available Dip-Former cleaning facilities. Use of 2.0 wt % of precipitated calcium carbonate is sufficient in many cases, if on-line acid cleaning can be carried out at least once every 12 circuits, and an alkaline chelating combination, such as sodium hydroxide 10.0 wt % with 3.0 wt % EDTA, is also used every circuit to remove residual organic matter from the formers.

Examples of nonionic surfactants are Emulphor ON and the Igepals (both Antara Chemicals), and Triton X-100 (Rohm & Haas Company). Release agents are fine, insoluble powders suspended in the coagulant and deposited on the form with the coagulant. Talc, clay, or diatomaceous earth can be used, but calcium carbonate is used more commonly because of ease of its removal from the form. Snow-Floss (Johns Mandeville Corporation) is also suitable, and in some coagulants also has a slight thickening effect, increasing amount deposited on the form. Of the three coagulant formulas shown, type *A* may be considered general-purpose and suitable for most dipping applications. Type *B* which contains the surfactant is used under conditions where the first type does not wet the forms evenly. Type *C* is used where the rubber tends to stick to the form because of its surface or shape, making removal of the finished goods difficult. The release agent also provides a layer of lubricant between the rubber and form, making the product's removal (stripping) easier.

In dipping generally, but particularly with the anode process, it is desirable to use tanks that circulate the coagulant and latex compound, particularly the latter. Use of circulation keeps the liquid surface clean and free from lumps, scum, or bubbles. Mechanical circulation can cause rubber particle instability, however, and eventually coagulate the compound. Therefore, tanks should be designed to minimize friction or shear action, and the compound stabilized to maintain mechanical stability.

Kaysam Process. The Kaysam process, which earlier had been used to some extent commercially, is similar to the roto-casting process with porous molds. However, in the Kaysam process the latex compound containing a gelling agent is introduced into metal molds and rotated until gelling takes place, usually with application of heat to accelerate gelling. The gelling agents are electrolytes with weak coagulating effect at room temperature, eg, solutions of ammonium salts and sodium silicafluoride with large amounts of zinc oxide or zinc carbonate. The process is based on the fact that latex sets to a gel with no change of volume, thereby taking on the shape of the mold, and as the gel dries it then shrinks uniformly and without distortion. Thus, by using a mold larger than the required size by an amount calculated from predetermined shrinkage, a satisfactory article can be made. Adequate washing of Kaysam molding is essential for good aging resistance.

Heat-Sensitizing Process. Another process used to make latex rubber articles of thicker section involves sensitizing the compound so that it coagulates when heated to a given temperature, then using heated molds to build the article to the desired thickness. Ammonia-preserved latex is used in this process, and polyether, polythioether, or poly(vinyl methyl ether) (PVME) can be used as heat-sensitizing agents.

Continuous Cast-Rubber Sheet. Gelling agents may also be used in continuous production of latex rubber sheeting, where wet latex compound is poured onto a flexible stainless steel or treated rubber conveyor belt, and its thickness controlled by a combination of passing the wet layer under a doctor blade and adjusting the speed of the belt as it does so. Application of heat causes thicker, heat-sensitized coatings to initially gel and then dry, prior to vulcanization. Thinner coatings can be formed and heat dried without need of a gelling agent.

Latex Rubber Foam. The flexible foam market, especially in the furniture and transportation industries, had earlier been overtaken by the flexible polyurethane foams because of their wide range of physical properties. The high resilience (hr) type especially shows superior properties and lower energy demand over the conventional molded hot-cure rubber foam. Latex foam, however, still constitutes an important application for natural and synthetic rubber latices, and demand for it has increased in the furniture industry because its combustion does not produce the toxic cyanate vapor generated by polyurethane-filled furniture. Natural rubber foam can also be compounded with fire-retardant properties. Latex foam is one of the most interesting latex applications and probably the most challenging from the viewpoint of completely understanding all the phenonema that occur during its manufacture.

Many different processes are patented for preparing latex foam but only two are of commercial interest for preparing molded cushioning stock: the Dunlop, which is most widely used, and the Talalay processes. Some producers have developed variations, which are combinations of the two processes.

The basic aspects of the two processes are whipping the latex to a froth by the mechanical incorporation of air into the latex, setting the frothed structure with a coagulant or gelling agent, and vulcanizing the rubber so that the structure becomes permanent. The proper coagulation of latex to give a stable foam (gelation) is the key to successful application. The gelling agent is one that can be mixed into the frothed latex and then remain dormant long enough to enable the froth to be poured into molds before exerting its gelling effect. Such a material is sodium silicofluoride, Na₂SiF₆. The actual coagulating system does not involve this compound alone; zinc oxide is believed to also play an active part. The pH of the aqueous phase begins to fall as soon as sodium silicofluoride is introduced, as a result of the progressive formation of hydrofluoric acid. The ZnO is hydrolyzed at ca pH 10. At pH 8–9, the zinc forms insoluble soaps with fatty acid latex stabilizers in the latex. The latex particles are sufficiently destabilized by removal of fatty acids that they

coalesce and form a gel. The pH continues to fall and ultimately reaches equilibrium between pH 7 and 8, and the bubbles in the system break to form an interconnecting cellular structure (55). The Si(OH)₄ may form in extremely fine particles, exposing a tremendous surface area which could adsorb the stabilizer and further act to induce gelation. Some latices may be so stable that gelation does not occur before foam collapse. Then so-called secondary gelling agents or sensitizers can be employed to destabilize the system sufficiently to cause it to gel before foam collapse. Cationic soaps, salts of *N,N'*-diphenylguanidine (DPG), and amine compounds known as Trimene Base are employed for this purpose; presumably these allow gelation to occur at a higher pH. DPG is also a rubber accelerator that is particularly effective when present in combination with a zinc dithiocarbamate.

The latex may consist entirely of natural latex or synthetic SBR latex or may be a mixture of both. In the Dunlop process, natural rubber foams shrink more than SBR foams during washing and drying. The load-bearing capacity of the foams at a given density falls significantly as SBR is used in place of natural rubber.

After the compound is prepared, it is whipped either on a batch or continuous basis. The Oakes continuous mixer is the standard piece of equipment in the industry. Then the gelling agents, ie, Na₂SiF₆ and ZnO, are added, the foam poured into molds and cured, and the product stripped from the mold for further cure.

In the Talalay process, the froth is produced by chemical rather than mechanical means. Hydrogen peroxide and an enzyme decomposition catalyst are mixed into the latex and the mixture placed in the mold. Decomposition of the peroxide by the added enzyme results in the liberation of oxygen which causes the latex mix to foam and fill the mold. The foam is then rapidly chilled and CO₂ is introduced to gel the latex. The gelled foam is then handled in a manner similar to that used in the Dunlop process.

For supported flat-stock foam, ie, foam backing on various fabrics such as carpets, scatter mats, upholstery fabrics, etc, either ammonium acetate or ammonium sulfate is employed in combination with zinc oxide as the gelation agent. The froth is prepared with an Oakes machine, the gelling agent is added at the machine, and the foam is applied to the fabric by spreading directly onto the fabric or spreading onto a belt and transferring the wet gel to the fabric. Gelling is carried out at elevated temperatures, usually with the aid of infrared lamps. To prevent uneven shrinkage, the fabric is carried through the high temperature zone and drying ovens on tenter frames. For this application the foam is poured in narrow thickness (3.2 to 12.7 mm).

Back Sizes. Backsizing defines the process of applying a film-forming adhesive material to the underside of carpets, rugs, or upholstery fabrics to achieve an anchor or lock of the pile yarns to the basic weave or backing fabric, firmer hand, increased weight, to overcome raveling of cut edges, antiskid properties, improved laying characteristics that are especially important in wall-to-wall carpeting, and improved dimensional stability. In the cases of tufted carpet and pile fabric, the pile anchoring or tuft-locking function of the back size is critically important. A latex that accepts a high loading of filler for low costs and at the same time develops the desired anchorage is required. The medium styrene–butadiene latices and particularly the so-called self-curing carboxy-modified SBR latices are most widely used.

The desired hand and drape is obtained by the proper selection of fillers and filler-loading ratio. Filler-loading ratio is the largest single factor in determining raw material costs. The higher the loading, of course, the lower the cost of the compound. A high ratio of binders-to-filler results in a strong backing of tufts or pile, improved dimensional stability, and excellent drape and nonskid properties. Common fillers are whitening and soft clays plus some titanium dioxide for opacity. The stability of the size to degradation and discoloration under the influence of light and heat is important.

Paper Applications. In beater additions, the latex is mixed with the beaten paper pulp either by addition at the beater or to the stock chest at the wet end of the paper machine. In either case, the pH of the pulp is reduced to 4.0–4.5, usually by the addition of a solution of alum to the pulp–latex mixture which has been thoroughly agitated. The latex, which for this application must be based on an anionic emulsifier, coagulates as the pH drops. The latex solids separate in intimate association with the pulp fibers. The pulp is then screened and the paper web formed in the conventional way. A latex for this purpose must possess the proper balance between mechanical and chemical stability.

At low latex solids-to-pulp ratios, ie, 10–20 pph, latex is added to the beaten pulp to give a paper web with superior web strength, elongation, bursting strength, internal bond, and tear strength. The nitrile latices and medium styrene–butadiene are commonly used as beater additions. In a similar manner, latex can be deposited on asbestos fibers. Such compositions are used as gaskets, linoleum bases, etc.

Latex solids can be incorporated into a paper web either on or off the paper machine. In on-machine saturation, the extracted and sometimes partially dried web is passed through the saturation bath as a step in the papermaking. In off-machine saturation, the dry finished web is saturated by a converter. In both on- and off-machine saturation, the web is passed through a bath of the latex. The pickup is controlled by the combination of bath strength, web speed, moisture content, and squeeze roll pressure. The pickup usually covers the same range as for beater additions. The heavy applications of resinous solids to paper are generally made by saturation. Again the nitrile and medium styrene–butadiene latices are used for saturation and for the same reasons as in the case of beater addition, ie, improved tear, tensile strength, and elongation.

At high (50–100 pph) saturation levels, with a near thermoplastic elastomer, a leather-like material results that can be embossed. The paper web used in pressure-sensitive tape is prepared by latex saturation in order to give it sufficient internal strength to release without delamination.

Because the latex solids in the saturation process are deposited in the structure of the paper web by drying, the colloidal system is not as critical as with beater addition. Nonionic and amphoteric surface-active materials can be effectively used in the latices. A low surface tension and small particle size are desirable features.

Both functional and decorative coatings can be applied to paper from latices. The aqueous dispersions can be used on conventional paper converting machinery which usually cannot handle hot melts and solvent coatings. The lack of fire hazard because of absence of solvents is an added advantage of the latex system.

If the coating is to be functional, the latex solids are generally film forming and are deposited on the surface of the web in a continuous film. If the function of the film is to provide oil or grease resistance, a medium-to-high nitrile latex or a plasticized poly(vinyl chloride) or poly(vinylidene chloride) latex may be used. The decorative coating applied to paper from latex may be a thin film of thermoplastic resin which, after application to the web, is cast against a mirror-finished heated roll. A high gloss paper stock results from this technique. The saturated vinyl and acrylic latices are used. More commonly the latex functions as the adhesive for a high clay loaded coating. Latices have to a large extent displaced natural materials, eg, casein and starch, long used as filler binders for coatings applied to paper and paperboard. The latices are superior to earlier materials because of low viscosity and heavy coating weights. This combination of properties is necessary to the operation of high speed paper coating machines. Further, the latices provide higher gloss, better ink holdout, more flexibility, and better water resistance.

Adhesives. Both natural and synthetic latices are used in the preparation of combining adhesives for various applications. Latex adhesives are used for bonding sheets of fabric, paper, or leather, either to themselves or to other materials. Essentially there are two types of latex adhesive: dry and wet combining. In the case of dry combining, both materials to be combined are individually coated and dried. While still tacky, they are pressed together as they pass through a set of rollers. Natural latex is generally used for this application because it has a greater tack than synthetic latices. Wet combining involves application of the adhesive to only one of the two materials being combined. After the adhesive has been applied to one material and is still wet, the second material is laminated to it as both pass over drying rolls. Styrene–butadiene latices are generally used for wet combining. Nitrile and neoprene latices are used when oil resistance is required. The most important single outlet, which accounts for two-thirds of the use of natural rubber latex adhesives, is in the footwear industry, eg, for soling, lasting, and laminating.

Tire-Cord Solutions. The latex dipping or coating of tire cord prior to building it into a tire carcass is a fabric adhesive application (see TIRE CORD). The adhesive bonds to the carcass rubber and textile must be stronger than the cohesive strength of the adhesive mass. In the case of rayon cord, a mixture of medium styrene–butadiene and butadiene–styrene–vinylpyridine latices is generally applied. Nylon requires an adhesive with such strong functional groups that vinylpyridine latices are almost universally used. Polyester tire cord is generally inert and more reactive chemicals must be added to the resorcinol–formaldehyde latex (RFL) system to promote adhesion to rubber. A predip solutioning of the polyester cord is made using 4,4'-methylene-bis(phenylcarbanilate) (Bonding Agent P-1, Uniroyal; Hylene MP, DuPont) and an epoxy curative (NERO10, Nagase Trading Company). The pretreated polyester cord goes through a second dip consisting of a RFL treatment where the latex is a vinylpyridine. In most cases, the tire cord solutioning bath includes a preformed, water-soluble resin of the resorcinol–formaldehyde type in mixture with the latex.

Vulcanization. Vulcanization of articles made from latex compounds generally follows the same principles and methods as in vulcanizing

articles made from dry rubber compounds. However, there are a number of important differences. The method of curing articles in a mold by means of a hot press is not used with latex articles. In latex technology, molding refers to the forming of a hollow article by deposition from latex within a female mold, or onto a male mold, formed to impart the desired shape and where necessary to create a pattern or surface texture at the rubber–mold interface. After drying the rubber deposit, moderately high temperatures can be employed for vulcanization but without the application of mechanical pressure. Natural rubber compounds commence reversion and become both weakened and sticky at temperatures significantly above 125°C in air.

Hot air, steam, and hot water vulcanization is widely used in the latex industry, and fluid-bed heat transfer and electronic microwave curing has also been used. Cross-linking by electron radiation has been experimentally used, but has not yet been developed commercially.

Fast, low temperature curing rubber compounds can be prepared by initial heat prevulcanization of the liquid latex and are marketed commercially (Revultex, Doverstrand Corporation). Rubber deposited from these often needs little more heat than that required to dry the deposit, to achieve optimum tensile strength and elongation. Such compounds are often used by small companies manufacturing thin-wall dipped medical latex products, such as examination gloves, as few compound preparation facilities are needed by the dipping company.

By comparison, temperatures as high as 150°C are often required for mold-enclosed hard natural rubber compounds, where mold plattens are directly heated by steam or electricity. Synthetic latex rubber compounds, however, can be vulcanized at temperatures higher than those for natural rubber; neoprene and acrylonitrile–butadiene can be vulcanized at as high as 135°C.

Prevulcanized compounds, which have undergone the necessary cross-linkage while in the liquid state, are becoming more popular in industry. Articles made from these can often be merely dried to be adequate for a purpose, but to achieve optimum physical properties it is usually necessary to give the articles a short, low temperature (ca 80°C) cure in an oven.

Prevulcanized natural rubber latices generally are used in making dipped goods or products produced by casting or rotational molding. The tear strength of prevulcanized latex rubber films can be inferior to those from post-vulcanized rubber deposits, but prevulcanization is widely used to manufacture compounds to make dip products ranging in wall thickness from 0.05 up to ca 3.0 mm.

Styrene–butadiene, acrylonitrile–butadiene, and butyl latices have also been prevulcanized, but this usually results in a lower wet gel strength. Where the rubber is to be deposited onto a fabric substrate, as in manufacture of fabric-lined gloves, the lower gel strength is not a problem. For unsupported dipped synthetic rubber products, however, the use of post-vulcanized compounds is more common.

Test Methods

Because of the long-range elasticity of soft rubber vulcanizates and the various special conditions under which they are degraded in use, special test methods have been developed which in many respects are unlike those used for wood, metals, and hard plastics. The American Society for Testing Materials (ASTM) through its Committee D11 on Rubber and Rubber-like Materials is constantly developing and promulgating new and improved methods for testing rubber and rubber products.

Unvulcanized Latex and Latex Compounds. A prime consideration has to be the fluid-state stability of the raw latex concentrate and liquid compound made from it. For many years, the mechanical stability of latex has been the fundamental test of this aspect. In testing, the raw latex rubber content is adjusted to 55% and an 80 g sample placed in the test vessel. The sample is then mechanically stirred at ultrahigh speed (ca 14,000 rpm) by a rotating disk, causing shear and particle collision. The time taken to cause creation of rubber particle agglomerates is measured, and expressed as the mechanical stability time (MST).

More recently, a number of tests of chemical stability of the latex concentrate have been developed. Chemical stability variance in the raw concentrate has considerable effect on the dipping characteristics of latex compounds, and can also affect mechanical stability of the compound. A broad rule is that, while latex MST can be increased or decreased without necessarily affecting its chemical stability, any change in the latter always is reflected in the MST. A new test, in which chemical stability is determined by measurement of the effect of weak zinc acetate solution added to a second mechanical stability sample and the result contrasted with the original MST, is available to numerically quantify chemical stability (56).

Latex compound viscosity obviously forms an important aspect of dipped product manufacture. Accurate measurement by a Brookfield or similar viscometer is desirable to establish the fundamental viscosity of a compound, but Flow-Cup viscometers (Ford B.3 Cup) are more commonly used for day-to-day control of latex compounds during compounding and product manufacture. It is necessary to ensure that only stainless steel flow cups are used, if the measured latex is allowed to return to the production tanks; brass cups yield an unacceptable level of copper contamination, which adversely affects aging properties of products made from copper-contaminated rubber compound.

For some dipped work, wet gel strength measurement is carried out, where the gel is deposited onto preformed, tissue paper dumbbells, and the load necessary to break the dumbbell-shaped test piece is measured.

A further test to establish extent of prevulcanization (cross-linkage) in a natural latex compound is based on its ability to withstand the swelling action of a mineral solvent. The test, termed the Swelling Index, involves forming a thin (ca 0.10 mm) even film of the wet latex on a glass plate and then drying it with a cool airstream; the airstream must not be strong enough to cause any ripple in the film. After drying, the film is dusted with talc or starch and peeled off the plate. A circular test disk of 40 mm dia is cut with a mechanical die-punch, and placed in toluene or another suitable solvent (SBP 2, Exxon Corporation) for 30 min. The film diameter is then measured and the new surface area compared with that of the original sample. As prevulcanization increases, the extent to which a sample swells decreases, and from that a scale or index of cross-linkage can be made.

Tests on raw or compounded but unvulcanized materials are chiefly concerned with rheological properties, that is, the response to the forces imposed during the operations of mixing, extrusion, calendering, and curing.

BIBLIOGRAPHY

"Rubber Compounding and Fabrication" in *ECT* 1st ed., Vol. 11, pp. 892–946, by A. E. Juve and E. G. Partridge, B. F. Goodrich; in *ECT* 2nd ed., Vol. 17, pp. 543–645, by R. R. Barnhart, Uniroyal Chemical; "Rubber Compounding" in *ECT* 3rd ed., Vol. 20, pp. 365–468, by R. R. Barnhart, Uniroyal Chemical.

1. *Worldwide Rubber Statistics*, International Institute of Synthetic Rubber Producers, Inc., 1994.
2. "Engineering Properties of Elastomers," *Plant. Eng.* (May 28, 1981).
3. *Synthetic Rubber End Use Survey*, International Institute of Synthetic Rubber Producers, Inc., 1990.
4. W. W. Barbin and M. B. Rodgers, *Science and Technology of Rubber*, Academic Press, Inc., New York, 1994, p. 424.
5. *Synthetic Rubber End Use Survey*, International Institute of Synthetic Rubber Producers, Inc., 1990.
6. F. R. Wolf and R. D. Demarco, *Vanderbilt Rubber Handbook*, R. T. Vanderbilt Co., Norwalk, Conn., 1978, p. 169.
7. J. R. Purdon, *Vanderbilt Rubber Handbook*, R. T. Vanderbilt Co., Norwalk, Conn., 1990, p. 167.
8. Ref. 6, p. 300.
9. Ref. 6, p. 188.
10. Ref. 6, p. 275.
11. A. Y. Coran, *Vulcanization*, Internal Technical Review, Monsanto Chemical Co., St. Louis, Mo., 1988.
12. R. S. Thompson, *Rubber in America Before 1492*, Godfrey Cabot Inc., Boston, Mass., 1950.
13. D. Shaw, "People Who Made It Possible," *Eur. Rubb. J.* **25** (June 1989).
14. W. Warrach, personal communication, Miles, Inc., Akron, Ohio, 1991.
15. Technical data, International Institute of Synthetic Rubber Producers, 1990.
16. Personal communication, Miles, Inc., Akron, Ohio, 1992; *Chem. Mark. Rep.* (Oct. 1991).
17. C. P. Rader, *Vulcanization, Intermediate Correspondence Course*, Rubber Division, American Chemical Society, Akron, Ohio, 1985.
18. R. W. Layer, *Elastomerics* (May 1988).

19. W. F. Helt, B. H. To, and W. W. Paris, "Post-Vulcanization Stabilization of NR," *Rubb. World*, **8** (1991).

20. J. A. Brydson, *Rubber Chemistry*, Applied Science Publishers, Ltd., London, 1978.

21. W. W. Paris, *Vulcanization, Its Activation and Acceleration*, Education Symposium Rubber Division, American Chemical Society, Akron, Ohio, Fall 1982.

22. M. Studabaker, *Vulcanization of Hydrocarbon Rubbers*, Phillips Chemical Co., 1970.

23. *Vulkealant E Retarder*, Product Information Bulletin, Miles, Inc., Akron, Ohio, 1987.

24. *Crosslinking Structures and Elastomers Properties*, Technical Bulletin, Miles, Inc., Akron, Ohio.

25. *Vulcanizing Nitrile Rubber*, Technical Bulletin HM-9, B. F. Goodrich Chemicals Co., Independence, Ohio.

26. *The Vanderbilt Rubber Handbook*, 13th ed., R. T. Vanderbilt Co., Norwalk, Conn., 1990.

27. *Vulcanizing Ethylene–Propylene Elastomers*, Technical Bulletin, No. ORC-104C, Hercules, Inc., Wilmington, Del.

28. G. P. Box, W. G. Hunter, and J. S. Hunger, *Statistics for Experimenters*, John Wiley & Sons, Inc., New York.

29. J. J. Luecken and M. A. Fath, "Rubber Chemicals Stability," *Kautschuk and Gummi Kunststoffe*, **35**(6) (1982).

30. *Nitrosamines in the Rubber Industry*, Technical Bulletin, Miles, Inc., Akron, Ohio.

31. R. F. Wolf, *Columbia Southern Chemical Corporation Laboratory Data*, Barberton, Ohio, Nov. 7, 1956.

32. J. R. Shelton, *Rubber Chem. Technol.* **45**, 359 (1972).

33. J. R. Shelton, in W. L. Hawkins, ed., *Polymer Stabilization*, John Wiley & Sons, Inc., New York, 1972.

34. L. D. Loan, R. W. Murray, and P. R. Story, *J. IRI* **2**, 73 (1968).

35. J. R. Dunn, *Rubber Chem. Technol.* **34**, 686 (1961).

36. B. M. Neyland and R. L. Stafford, *Trans. IRI* **35**, 25 (1959).

37. M. A. Plant and G. Scott, *Eur. Polym.* **7**, 1173 (1971).

38. R. B. Spacht, W. S. Hollinshead, H. L. Bullard, and D. C. Willis, *Rubber Chem. Technol.* **38**, 134 (1965).

39. H. Long, ed., *Basic Compounding and Processing of Rubber*, 1985, p. 28.

40. E. J. Latos and A. K. Sparks, *Rubber J.* **151**(6), 18 (1969).

41. J. A. Kuczkowski and J. G. Gillick, *Rubber Chem. Technol.* **57**(3), 621 (1984).

42. R. O. Babbitt, *Vanderbilt Rubber Handbook*, R. T. Vanderbilt Co., Norwalk, Conn., 1978, p. 564.

43. G. S. Whitby, ed., *Synthetic Rubber*, John Wiley & Sons, Inc., New York, 1954.

44. F. Bovey and co-workers, *Emulsion Polymerization*, Vol. 10, *High Polymers*, Interscience Publishers, New York, 1955.

45. L. H. Howland and co-workers, *Rubber Plast. Age* **42**, 868 (1961).

46. L. Talalay, *Rubber Chem. Technol.* **36**(3), 581 (July–Sept. 1963).

47. R. B. Spacht and co-workers, *Rubber Chem. Technol.* **37**(1), 210 (Jan.–Mar. 1964).

48. R. O. Becker and K. L. Seligman, *Rubber Chem. Technol.* **34**(3), 856 (July–Sept. 1961).

49. E. L. Borg, *Rubber World* **137**, 723 (1958).

50. H. P. Brown, *Rubber Chem. Technol.* **30**(5), 1382 (Dec. 1957).

51. H. C. Jones and C. A. Klamann, *Rubber Age* **70**, 325 (1951).

52. **U.S. Pat. 1,548,689** (Aug. 4, 1925), P. Klein (to Anode Rubber Co., Ltd.).

53. **U.S. Pat. 1,996,051** (Apr. 2, 1935), D. F. Twiss (to American Anode, Inc.).

54. **U.S. Pat. 1,719,633** (July 2, 1929), M. C. Teague (to U.S. Rubber Co. (Uniroyal)).

55. F. A. Murphy, *Trans. Inst. Rubber Ind.* **31**, 105 (1955).

56. *ACS Spring Meeting*, American Chemical Society, Rubber Division, Washington, D.C., 1995.

General References

Materials Science

J. C. Anderson, K. D. Leaver, J. M. Alexander, and R. D. Rawlings, *Materials Science*, 2nd ed., Halsted Press, a division of John Wiley & Sons., Inc., New York, 1974.

K. P. Beardsley, "An Introduction to the Physics of Material Science," at the *ACS Meeting of the 147th Rubber Division, Philadelphia, Pa.*, May 2–5, 1995, American Chemical Society, Washington, D.C.

K. L. Watson, *Materials in Chemical Perspective*, Halston Press, a division of John Wiley & Sons, Inc., New York, 1975.

General

A. K. Bhowmick, M. M. Hall, and H. A. Benarey, *Rubber Products Manufacturing Technology*, Marcel Dekker, Inc., New York, 1994.

Designing With Elastomers, Rubber Energy Group.

Glossary of Terms Relating to Rubber and Rubber Technology, American Society for Testing and Materials, Philadelphia, Pa., 1972.

J. E. Mark, B. Erman, and F. R. Eirich, *Science and Technology of Rubber*, 2nd ed., Academic Press, Inc., New York, 1994.

M. Morton, *Introduction to Rubber Technology*, Reinhold Publishing Corp., New York, 1959.

M. Morton, ed., *Rubber Technology*, Robert E. Krieger Publishing Co., Malabar, Fla., 1981.

M. Morton, ed., *Rubber Technology*, 3rd ed., Van Nostrand Reinhold Co., New York, 1987.

K. F. O'Driscoll, *The Nature and Chemistry of High Polymers*, Reinhold Publishing Corp., New York, 1964.

R. F. Ohm, ed., *The Vanderbilt Rubber Handbook* 13th ed., R. T. Vanderbilt Co., Inc., Norwalk, Conn., 1990.

Rubbicana Rubber Directory and Buyer's Guide, Crain Communications, Inc., 1994.

D. R. Smith, *Blue Book 1995*, Lippincott & Peto, Inc., Akron, Ohio, 1995, pp. 384–387.

R. C. Weast, ed., *CRC Handbook of Chemistry and Physics*, 1st student ed., CRC Press, Inc., Boca Raton, Fla., 1988.

Polymer Science

F. W. Billmeyer, Jr., *Textbook of Polymer Science*, Interscience Publishers, a division of John Wiley & Sons, Inc., New York, 1962.

J. M. Charrier, *Polymeric Materials and Processing*, Hanser Publishers, New York, 1990.

R. Mazzeo, *Polymer Degradation and Its Prevention Especially During Mixing*, at the *ACS Meeting of the 147th Rubber Division, Philadelphia, Pa.*, May 2–5, 1995, American Chemical Society, Washington, D.C.

D. W. Van Krevelen, *Properties of Polymers*, 3rd ed., Elsevier Scientific Publishing Co., Amsterdam, the Netherlands, 1990.

A. V. Tobolsky and H. F. Mark, *Polymer Science and Materials*, Wiley-Interscience, a division of John Wiley & Sons, Inc., New York, 1971.

Elastomers

AFLAS Processing Guidelines, 3rd ed., Xenox, Inc., 1985.

L. D. Albin and M. M. Lynn, *Chemistry of Elastomers Containing Vinylidene Fluoride*, at the *ACS Meeting of the Energy Rubber Group, Rubber Division, Educational Symposium, Arlington, Tex., Sept. 23, 1985*, American Chemical Society, Washington, D.C.

A. K. Bhowmick and H. L. Stephens, *Handbook of Elastomers*, Marcel Dekker, Inc., New York, 1988.

Bromobutyl Handbook, Technical Publications Section, Technical Development Division, Polysar Ltd., Sarnia, Canada.

Bromobutyl Rubber Compounding and Applications, Exxon Chemical Co.

N. L. Catton, *The Neoprenes*, E. I. du Pont de Nemours and Co., Inc., Wilmington, Del., 1953.

J. R. Dunn and G. C. Blackshaw, *NBR—Chemistry and Markets*, at the *ACS Energy Rubber Group Educational Symposium, Dallas, Tex., Sept. 24, 1985*, American Chemical Society, Washington, D.C.

Fhuorel/Fhuorelastomer, 3M Industrial Chemical Products Division.

D. L. Hertz, *Chemistry of AFLAS*, at the *ACS ERG-Education Program, Arlington, Tex., Sept. 24, 1985*, American Chemical Society, Washington, D.C.

R. F. Karg, *EPDM Polymer Chemistry Review*, presented to West Michigan Rubber Group, Cadillac, Mich., Sept. 7, 1994.

D. R. Matheson, *Nitrile Compounding Symposium*, presented at Rubber Energy Group, Sept. 1985.

R. M. Murray and D. C. Thompson, *The Neoprenes*, E. I. du Pont de Nemours and Co., Inc., Wilmington, Del., 1963.

Polysar Handbook, Vol. 1, Polymer Corp. Ltd., Sarnia, Canada, 1956.

Polysar Handbook, Vol. 2, Polymer Corp. Ltd., Sarnia, Canada, 1960.

K. M. Pruett, *Chemical Resistance Guide for Elastomers*, Compass Publications, 1988.

H. E. Rooney, ed., *Polysar Butyl Handbook*, Polymer Corp. Ltd., Sarnia, Canada, 1966.

J. H. Saunders and K. C. Frisch, *Polyurethanes Chemistry and Technology*, Part 1, *Chemistry*, Interscience Publishers, a division of John Wiley & Sons, Inc., New York, 1962.

J. H. Saunders and K. C. Frisch, *Polyurethanes Chemistry and Technology*, Part 2, *Technology*, Interscience Publishers, a division of John Wiley & Sons, Inc., New York, 1964.

Synthetic Rubber End-Use Survey, International Institute of Synthetic Rubber Producers, Inc., 1990.

The Synthetic Rubber Manual, 12th ed., International Institute of Synthetic Rubber Producers, Inc., 1992.

D. C. Thompson, *Mechanical Molded Goods, Neoprene and Hypalon*, E. I. du Pont de Nemours and Co., Inc., Wilmington, Del., 1955.

Vistalon Ethylene-Propylene Rubber User's Guide, Exxon Chemical Co., Polymers Group.

B. M. Walker and C. P. Rader, ed., *Handbook of Thermoplastic Elastomers*, 2nd ed., Van Nostrand Reinhold Co., New York, 1988.

W. C. Warner, *Methods of Devulcanization, Rubber Chemistry and Technology, Rubber Reviews*, Rubber Division, American Chemical Society, Inc., Washington, D.C., 1994, pp. 559–566.

G. S. Whitby, ed., *Synthetic Rubber*, John Wiley & Sons, Inc., New York, 1954.

K. M. Wolniak, "Butyls and Halobutyls," presented to the *Akron Rubber Group Educational Lecture Series*, Akron, Ohio, Mar. 13, 1989.

Worldwide Rubber Statistics 1994, International Institute of Synthetic Rubber Producers, Inc., 1994.

Zetpol Technical Note, Zeon Chemicals, Inc.

Vulcanization

G. Alliger and I. J. Sjothun, *Vulcanization of Elastomers*, Reinhold Publishing Corp., New York, 1964.

W. H. Helt and D. Sikora, *Accelerated Sulfur Vulcanization*, at the *ACS Meeting of the 147th Rubber Division, Philadelphia, Pa., May 2–5, 1995*, American Chemical Society, Washington, D.C.

W. Hofmann, *Vulcanization and Vulcanizing Agents*, Maclaren and Sons Ltd., London, Palmerton Publishing Co., Inc., New York, 1967.

Fillers

J. Byers, *Reinforcing Carbon Blacks*, at the *ACS Meeting of the 147th Rubber Division, Philadelphia, Pa., May 2–5, 1995*, American Chemical Society, Washington, D.C.

J. Dacey and W. Fultz, *Synthetic Silicate Products for the Rubber Industry*, at the *ACS Meeting of the 147th Rubber Division, Philadelphia, Pa., May 2–5, 1995*, American Chemical Society, Washington, D.C.

N. Hewitt, *An Introduction to Silica Fillers*, at the *ACS Meeting of the 147th Rubber Division, Philadelphia, Pa., May 2–5, 1995*, American Chemical Society, Washington, D.C.

Hi-Sil Silica Products Formulary, Vol. 3, PPG.

J. L. Jones and S. B. Radding, *Thermal Conversion of Solid Wastes and Biomass*, American Chemical Society, Washington, D.C., 1980.

S. Monthey and X. Zhang, *Carbon Black Selection for Industrial Rubber Applications*, at the *ACS Meeting of the 147th Rubber Division, Philadelphia, Pa., May 2–5, 1995*, American Chemical Society, Washington, D.C.

O. Noel, *Talk, A Functional Mineral for Rubber*, at the *ACS Meeting of the 147th Rubber Division, Philadelphia, Pa., May 2–5, 1995*, American Chemical Society, Washington, D.C.

Product Application in the Rubber Industry, DeGussa.

D. R. Schultz, *Flame Retardants—Specialty Fillers*, at the *ACS Meeting of the 147th Rubber Division, Philadelphia, Pa., May 2–5, 1995*, American Chemical Society, Washington, D.C.

M. Tapper, *Calcium Carbonate Fillers for Elastomers*, at the *ACS Meeting of the 147th Rubber Division, Philadelphia, Pa., May 2–5, 1995*, American Chemical Society, Washington, D.C.

F. J. Washabaugh, *Kaolins for Rubber Applications*, at the *ACS Meeting of the 147th Rubber Division, Philadelphia, Pa., May 2–5, 1995*, American Chemical Society, Washington, D.C.

Oils and Plasticizers

G. L. Brown, *Resins in Rubber*, Pennsylvania Industrial Chemical Corp., Clairton, Pa., 1969.

N. A. J. Platzer, *Plasticization and Plasticizer Processes*, American Chemical Society, Washington, D.C., 1965.

General Compounding

P. J. Corish, *Concise Encyclopedia of Polymer Processing and Applications*, Pergamon Press, New York, 1992.

P. R. Dean II, *Index of Commercial Antioxidants & Antiozonants*, 4th ed., Goodyear Chemicals, Akron, Ohio, 1983.

J. Edenbaum, ed., *Plastics Additives and Modifiers Handbook*, Van Nostrand Reinhold, New York, 1992.

Harwick Chemical Corporation Custom Materials for Rubber and Plastics, Harwick Chemical Corp.

Materials for Rubber and Plastics Products, Akrochem Corp., 1989.

Strucktol Processing Agents, Struktol.

Latex

J. C. Carl, *Neoprene Latex*, E. I. du Pont de Nemours and Co., Inc., Wilmington, Del., 1962.

Testing

Annual Book of ASTM Standards, Index, Vol. 00.01, American Society for Testing and Materials, Philadelphia, Pa., 1994.

Annual Book of ASTM Standards, Rubber, Vol. 09.01, American Society for Testing and Materials, Philadelphia, Pa., 1994.

Annual Book of ASTM Standards, Rubber, Vol. 09.02, American Society for Testing and Materials, Philadelphia, Pa., 1994.

ASTM Standards on Carbon Black, American Society for Testing and Materials, Philadelphia, Pa., 1991.

W. Bigot, *Dynamic Measurement Practices of Elastomeric Materials and Components*, presented at the *ACS Meeting of the 147th Rubber Division, Philadelphia, Pa., May 2–5, 1995*, American Chemical Society, Washington, D.C.

M. Roland, *Viscoelasticity—Material Behavior and Measurement Technique*, at the *ACS Meeting of the 147th Rubber Division, Philadelphia, Pa., May 2–5, 1995*, American Chemical Society, Washington, D.C.

H. Witt, *Overview of the Finite Element Method with Emphasis on Applications to Rubber Materials*, at the *ACS Meeting of the 147th Rubber Division, Philadelphia, Pa., May 2–5, 1995*, American Chemical Society, Washington, D.C.

William Klingensmith
Frank Jenkins
Akron Consulting Company

Walter Klamp
Consultant

Michael Fath
Consultant

Kris Baranwal
ARDL

Robert Ohm
R. T. Vanderbilt Company

Ray Russel Fell
Consultant

Brendon Rodgers
Consultant



RUBBER, NATURAL

Virtually all the natural rubber [9006-04-6] EINECS 23 26890, used as of 1996, is obtained from the latex of the *Hevea brasiliensis* tree. The rubber tree grows best in a warm, damp, uniform climate, with temperatures remaining in the 24–28°C range year-round, humidity above 70% , and a well-distributed rainfall of 1800–2000 mm yr falling on well-drained soils. It is usually limited to areas extending 15° on either side of the equator, where it can survive up to an altitude of 300 m. Rubber trees were originally found in South America, particularly in the Amazon valley, and in 1876 Sir Henry Wickham collected some 70,000 seeds at the request of the India office and brought them to Royal Botanic Gardens at Kew, in England. About 2000 of the seeds were successfully cultivated into plants, some of which were sent to Peradeniya in Sri Lanka and to Malaysia via the Singapore Botanical Gardens (1). Later shipments were sent to Indonesia, and by 1880 *Hevea* seedlings were widely distributed in Southeast Asia.

The rubber tree quickly flourished in Malaysia, where large areas of jungle were cut down and planted with rubber trees. Henry Ridley, who was appointed Director of the Singapore Botanical Gardens in 1888, did perhaps more than anyone else to encourage the planting of rubber, and in the 1890s devised modern methods of tapping that resulted in the economic harvesting of the rubber crop (2,3). Further development, by John Parkins, of the coagulation process in which dilute acid was added to the latex also helped to develop the economic viability of the plantation industry.

By the end of the nineteenth century, there were 2500 ha of land under rubber cultivation in Southeast Asia. The start of the automotive era shortly afterward caused a rocketing demand for rubber, and by 1910 there were 500,000 ha of rubber planted and the countries of Southeast Asia had become the main suppliers of natural rubber worldwide.

It is thought that only 24 seedlings survived the original journey to Malaysia (4), and such a small number, coupled with selective breeding for high yield, has produced a narrow genetic base. In order to widen the base, the International Rubber Research and Development Board, in collaboration with Brazilian authorities, carried out a germplasm prospection in the Brazilian jungles in 1981.

Rubber Production

Dry Rubber. All commercial grades of dry rubber arise from the organization of tapping and tappers and from the methods of coagulation and processing. Rubber production is generally managed in three ways: (1) the tapper can be a smallholder with 2–20 ha who will sell the crop, either in latex or semiprocessed form; (2) the tapper can be an employee of a self-contained estate or plantation of several thousand hectares with the whole operation of dry rubber production managed by the estate; or (3) the tapper can be part of a government-sponsored cooperative and will deliver the crop to central manufacturing stations.

The first cultivated rubber trees produced 500 kg ha of rubber and selective breeding over the years has resulted in yields during the mid-1990s as high as 3000 kg ha. However, average yields in the principal producing countries can vary from 400–1200 kg ha. In Malaysia the smallholder sector produces about 850 kg ha, whereas the estate plantations can produce about 1200 kg ha.

Latex and Its Collection. The latex from the *Hevea brasiliensis* tree is a colloidal dispersion consisting of nonrubber substances and rubber particles in an aqueous serum phase. The rubber hydrocarbon constitutes 30–45% of the whole latex; the nonrubber substances account for 3–5% . Rubber particles vary in size from 0.15–3 μ and contain 90–95% natural rubber, ie, *cis*-1,4-polyisoprene, with a molecular weight distribution of 10⁵–10⁷ g mol. The shape of the molecular weight distribution curve, which varies from a skewed unimodal to a bimodal distribution, and the molecular weight of natural rubber is determined by clone (5,6). In natural rubber, latex coagulation takes place under acid conditions at about pH 5. Coagulation can either be induced by the deliberate addition of acid or occurs naturally in the tapping cup by reaction with acids produced by bacterial action, ie, autocoagulation.

Rubber latex is formed and stored in rings of latex vessels found in the soft bark region of the tree, which lies between the inner cambium tissue and the outer hard bark layers. The physiological role of latex in the rubber tree is undetermined, as of 1996; however, the mechanism of production and methods of tapping are well understood (7,8). Latex is collected from the rubber tree by tapping once every two or three days. A tree is tapped by carefully cutting into the trunk with a special knife that removes a thin slice of bark and severs the latex vessels without affecting the normal sap circulatory system. The cut is made at an angle of 25–30° halfway around the circumference of the trunk, and at the lowest point a short metal spout is inserted, from which the latex flows into cups and is collected four to five hours later. It is also possible to apply a stimulant such as Ethephon (9), the active ingredient of which is 2-chloroethanephosphonic acid; the latter releases ethylene gas to increase the time during which latex flows and thus increase yield with a reduced tapping frequency. Unless the latex is to be processed immediately, a small amount of preservative, eg, ammonia, is added to prevent premature coagulation before the latex is brought to a factory or smallholder processing center. Rubber collected in this manner is known as field latex.

Latex continues to drip after the initial collection and coagulates naturally in the cup to form cup lump. Coagulum which forms as a film of latex on the tapped cut, called tree lace, or from latex that has dripped onto the ground, called earth scrap, is collected the next day along with the cup lump. Some smallholders may not collect the latex at all, but allow it to coagulate in the cup and collect it as cup lump. All these methods produce rubber known as field coagulum.

GRADE PRODUCTION OF DRY NATURAL RUBBER

Field latex and field coagulum are the source materials for all varieties and grades of dry natural rubber that include the conventional International grades as well as the Technically Specified Rubbers (TSR).

Conventional Grades. Classification of traditional grades such as ribbed smoked sheet (RSS) is based on the source of raw materials and processing methods. Visual grading and comparison with international samples depend on factors such as color, cleanliness, and uniformity of appearance as described in the "Green Book," which lists all 35 grades (10). A schematic diagram of the routes by which conventional grades are produced is shown in Figure 1. Of all the 35 traditional grades, only ribbed smoked sheet and pale crepes are traded in any substantial quantity, because cup lump materials are used to make the lower grades of Technically Specified Rubbers.

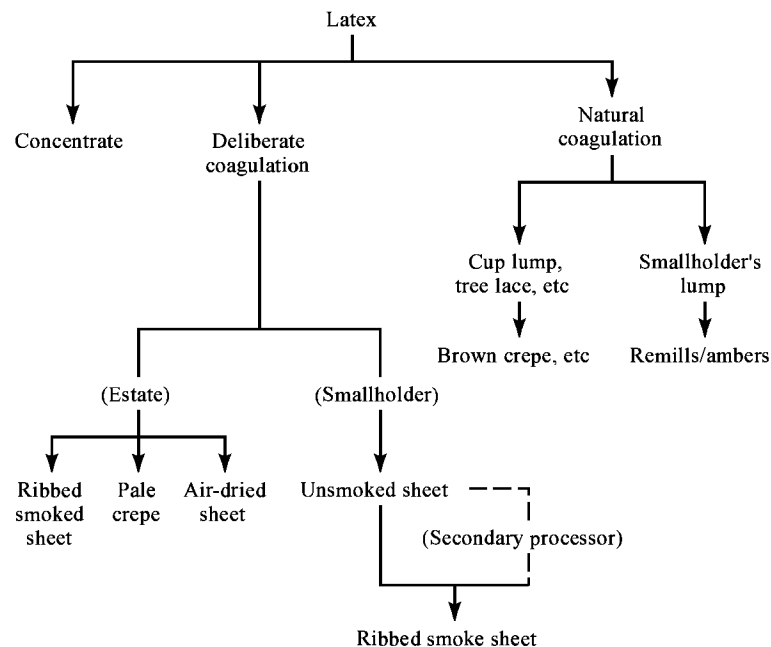


Fig. 1. Routes by which natural rubber latex is converted into traditional sheet and crepe.

Sheet Rubber. In large-scale operations, field latex is first preserved with either ammonia, sodium sulfite, or formalin and then diluted to ~15% dry rubber content (DRC) and strained into coagulating tanks. Formic acid (5 wt %) is added and the acidified latex left to stand for several hours in tanks to coagulate. Separators are positioned in the tank so that after coagulation the spongy coagulum is in the form of a continuous sheet. Water is squeezed from the coagulum by passing it through as many as six pairs of rollers to produce a thin sheet. Grooves on the last pair of rollers introduce the characteristic criss-cross rib markings on the sheet to aid drying by increasing the rubber surface area. Wet sheets are dried on trolleys in a smokehouse at 60°C over a period of about a week, and the resulting ribbed smoked sheets (RSS) are pressed into 112-kg bales. The bales are coated with a talc slurry to prevent bales sticking to each other during transport.

Pale Crepes. Sri Lanka is the largest producer of crepe rubber, as either thin pale crepe or thick pale crepe, which accounts for 35% of rubber production. Pale crepes are divided into one of three classes, depending on the manufacturing process (11,12): fractionated and bleached rubber (FB); unfractionated, but bleached rubber (UFB); or yellow fraction rubber (YF).

Light-colored pale crepes are made by selecting clonal species that give latices with low levels of naturally occurring yellow pigments (carotenoids) and so do not darken as much as other grades. However, sodium bisulfite is added to reduce this enzymatic darkening. Also, fractional coagulation to remove a pigment-rich fraction may be used in conjunction with a bleaching process to produce a light-colored latex. Latex is diluted to 20% DRC, followed by acid coagulation. The coagulum is milled and thoroughly washed by passing it through grooved rollers eight to nine times to produce thin crepes 1–2 mm thick. Thick pale crepe, or sole crepe, is made by plying thin crepes together to form a laminate about 1 cm thick, the surface of which is usually ribbed. The crepes are warm-air dried at 40°C over two weeks and packed as talc-coated 102-kg bales or polyethylene wrapped 33–35-kg bales.

Brown and Blanket Crepes. Several grades of these types of crepe are produced, and the source materials are field coagulum, unsmoked sheets, and smoked sheet cuttings. Such raw materials are thoroughly washed and blended through a series of crepers to produce long strips which are dried for two weeks in naturally ventilated buildings or by heating. When dry, the crepes are packed into talc-coated 102-kg bales.

Technically Specified Grades. The introduction of the Standard Malaysian Rubber (SMR) scheme in 1965 marked a turning point in the whole approach of production and marketing of natural rubber. The scheme was so successful that it led to the adoption of an international scheme for Technically Specified Rubber (TSR). These schemes led to fewer grades, all with guaranteed specifications relating to quality, packed in small, polyethylene-wrapped 33.3-kg bales for easy transport, storage, and factory handling. Faster drying processes for crumb rubber improved the efficiency of producing factories, and the technical specifications gave consumers a guarantee of quality that had not been previously available for the conventional grades.

A significant revision to the SMR scheme was introduced in 1991 in response to consumer desire for greater consistency in natural rubber (13). Other producing countries have similar specification schemes (14), as does the International Standards Organization (ISO) (15). An example of the specifications for TSR is given in Table 1 for the present Standard Malaysian scheme. Except for SMR 5, rheograph and cure test data (delta torque, optimum cure time, and scorch) are provided.

Table 1. Standard Malaysian Rubber (SMR) Specifications Scheme^a

Parameter	Latex			Sheet material	Blend	Field-grade material	
	SMR CV60	SMR CV50	SMR L	SMR 5 ^b	SMR GP	SMR 10	SMR 20
dirt retained on 44 μ aperture, max, wt %	0.02	0.02	0.02	0.05	0.08	0.08	0.16
ash content, max, wt %	0.50	0.50	0.50	0.60	0.75	0.75	1.00
nitrogen, max, wt %	0.60	0.60	0.60	0.60	0.60	0.60	0.60
volatile matter, max, wt %	0.80	0.80	0.80	0.80	0.80	0.80	0.80
Wallace rapid plasticity, P ₀ , min			35	30		30	30
plasticity retention index (PRI), min, %	60	60	60	60	50	50	40
Mooney viscosity, ML (1 + 4), 100°C	60 (±5)	50 (±5)			65 (±7)	d	d
color coding marker	black	black	light green	light green	blue	brown ^e	red ^f
plastic strip color ^g	orange	orange	transparent	opaque white	opaque white	opaque white	opaque white

^a Mandatory from Oct. 1, 1991.

^b Two subgrades of SMR 5 are SMR 5RSS and SMR 5ADS, which are prepared by direct baling of ribbed smoked sheet and air-dried sheet (ADS), respectively.

^c Special producer limits and related controls are also imposed by the Rubber Research Institute of Malaysia (RRIM) to provide an additional safeguard.

^d The Mooney viscosities of SMR 10CV and SMR 20CV (not listed) are not of specification status (ca 1996). They are, however, controlled at the producer end to 60 (+7, -5) for SMR 10CV and 65 (+7, -5) for SMR 20CV.

^e Magenta for SMR 10CV.

^f Yellow for SMR 20CV.

^g Plastic wrap color is transparent for all grades.

Production of TSR is different from conventional rubber production, and Figure 2 shows the relationship between source materials and SMR grades; other producing countries have similar methods of production. The conversion of liquid latex to the SMR latex grades is a large-scale factory operation, involving the use of heavy-duty processing equipment. The incoming latex is bulked and blended and formic acid is added to reduce the pH to 5.0–5.2 and to produce rapid coagulation. The coagulum is covered with water and allowed to mature for about 16 h. Dewatering, cracking, and comminution of the coagulum is obtained by hammer mills, granulators, shredders, and pelletizers and produces small, uniformly sized crumbs. There are two processes for crumbing natural rubber, the Heveacrub process where castor oil is used and the comminution process where no castor oil is used. In the Heveacrub process, 0.3% castor oil is added before coagulation to facilitate the process of comminution to a crumb form, and any excess oil is removed in subsequent washing operations. In all operations the final wet crumbs are dried in deep bed or conveyor driers at 100–120°C for four to five hours, followed by cooling, baling, and wrapping.

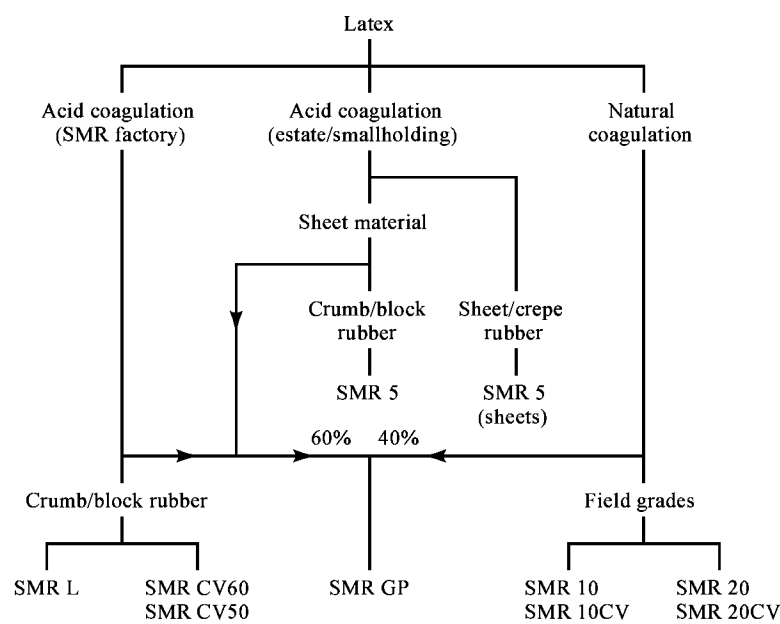


Fig. 2. Source materials and production processes used to make the various grades of Standard Malaysian Rubber (SMR).

Field coagulum is also converted to crumb form by similar size-reducing processes, but first it is important to wash the field coagulum thoroughly to remove bark and other contamination. The Heveacrub process may also be used in the production of field-grade rubbers. Figure 3 shows two schematic diagrams illustrating typical processes involved in the production of latex and field-grade rubbers.

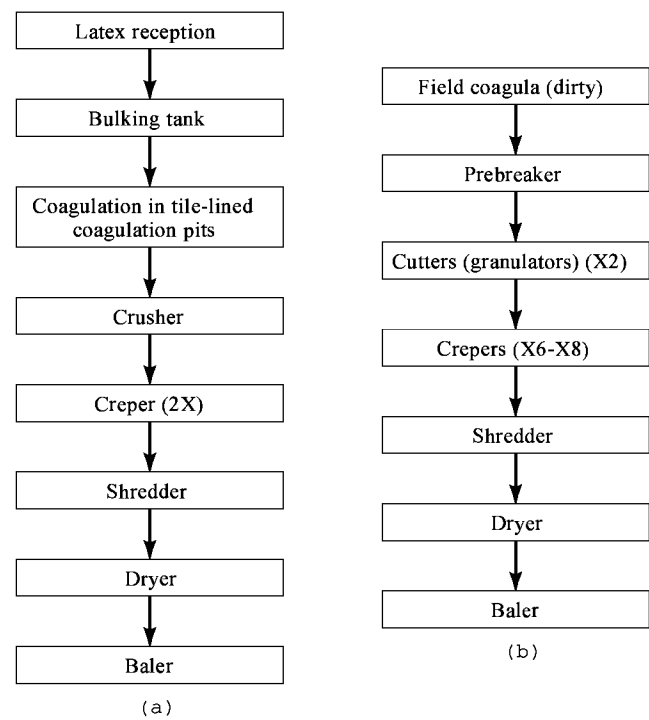


Fig. 3. Typical factory processing lines for (a) latex-grade rubber, and (b) field-grade rubber.

TSR L. TSR L is a color-specified rubber, and the light amber color is produced by selecting clones with a low carotenoid content. After collection, the field latex is preserved with a mixture of ammonia and boric acid and subsequently treated using 0.05% sodium metabisulfite to inhibit

enzymic darkening. The latex is coagulated by addition of formic acid and allowed to mature for up to 12 h. The resulting coagulum is processed into crumb form, followed by hot air drying at 100°C for roughly five hours. The dried crumbs are cooled to 60°C and compressed into standard bales of 33.3 kg and wrapped in polyethylene.

The color of TSR L is measured with a Lovibond Comparator using methods outlined in the relevant International Standard (16). Raw rubber is molded into a disk 1.6 mm thick, and the color is compared and matched with that of standard glass disks. These glass disks provide a color index scale in which the higher index values correspond to darker colors. Generally the specification for TSR L is a maximum Lovibond individual value of 6 and range (max) 2.

TSR CV. Natural rubber undergoes an irreversible increase in viscosity which occurs during transport and storage. This storage-hardening can be effectively inhibited by the addition of 0.15% hydroxylamine neutral sulfate (HNS) to ammonia-preserved latex prior to acid coagulation. The coagulum is then processed normally to produce TSR CV60 and TSR CV50 grades which have guaranteed Mooney viscosity ranges of 60 ± 5 and 50 ± 5, respectively.

TSR 5. Within the Malaysian scheme SMR 5 is restricted to rubber derived from sheet material prepared by conventional processes, ie, ribbed smoked sheet (RSS), air-dried sheet (ADS), and unsmoked sheet (USS) and presented in small bale form. Other countries may source different raw materials, eg, Indonesia produces SIR 5 from thin latex coagulum (14). SMR 5 prepared by pressing dry-sheet material into standard bales must be identified by the type of sheet material, eg, SMR 5RSS, on the test certificate as well as on the wrapping.

SMR GP. SMR GP is a general-purpose (GP) grade of viscosity-stabilized rubber made using a 60:40 mixture of latex-grade rubber sheet material and field coagulum. Its specifications are similar to SMR 10, with the additional guaranteed specification of Mooney viscosity (65 ± 7 5). Not less than 20% (dry weight) must be latex or sheet material of SMR 5 quality, processed in an SMR factory or designated processing center. The balance of the sheet material may be USS or RSS. SMR GP can be produced by two methods, the wet process and the dry process. In the wet process, cleaned crumbs of field coagulum are mixed with latex containing HNS and coagulated by the addition of formic acid. The mixture is then processed in the normal way to produce crumb rubber. The dry process involves the application of HNS as a concentrated spray onto mixed dried latex and field-grade rubber, which are then blended in an extruder, followed by cooling and baling.

TSR 10 and TSR 20. These field grades consist predominantly of cup lump or field coagulum, where natural coagulation of the latex has been allowed to occur in the collecting cups. The raw materials are first soaked in water, blended, then fed through a series of creper-hammermills, granulators, and shredders (methods of size reduction depend on each factory) to produce clean crumbs which are dried, then baled. Specification testing of the final rubber is carried out to determine which grade is produced and cleaner rubbers usually end up as the TSR 10 grade. Knowledge of the source of raw material and experience of processing also help to determine each grade.

TSR 10CV and TSR 20CV. Certain producing countries make viscosity-stabilized grades (controlled viscosity, CV) from field coagulum. In recognition of consumer demand for these particular grades, Malaysia introduced SMR 10CV and SMR 20CV into the revised SMR scheme in 1991. Processing is carried out using the same methods and raw materials as for other nonstabilized field grades. Viscosity stabilization is achieved by spraying HNS solution onto dried crumb before passage through a prebreaker. The Mooney viscosities of SMR 10CV and SMR 20CV are at present not of specification status. They are, however, controlled by the producer to 60 (±7, 5) and 65 (±7, 5), respectively (see Table 1).

Properties of Natural Rubber

Crystallization. Raw natural rubber may freeze or crystallize during transit or prolonged storage, particularly at subzero temperatures. The rubber then becomes hard, inelastic, and usually much paler in color. This phenomenon is reversible and must be differentiated from storage hardening. The rate of crystallization is temperature-dependent and is most rapid at -26°C. Once at this temperature, natural rubber attains its maximum crystallinity within hours, and this maximum is no more than 30% of the total rubber.

During crystallization, rubber molecules associate into a crystal array. As the storage temperature is increased, thermal motion of the molecules restricts crystallization until, at temperatures above 20°C, the rate is near zero. Observed melting temperatures of natural rubber depend markedly on the temperature at which crystallization occurs. Figure 4 shows a general trend in which the higher crystallization temperatures result in higher crystal melting temperatures (17). The wide range of melting temperatures exhibited by rubber crystallized at low temperature is a reflection of a polycrystalline texture that contains a large number of small crystals. The generally accepted crystal melting temperature of natural rubber is 30°C. Crystals with high melting temperature are fewer and larger than those formed at lower temperatures and take a relatively long time to form.

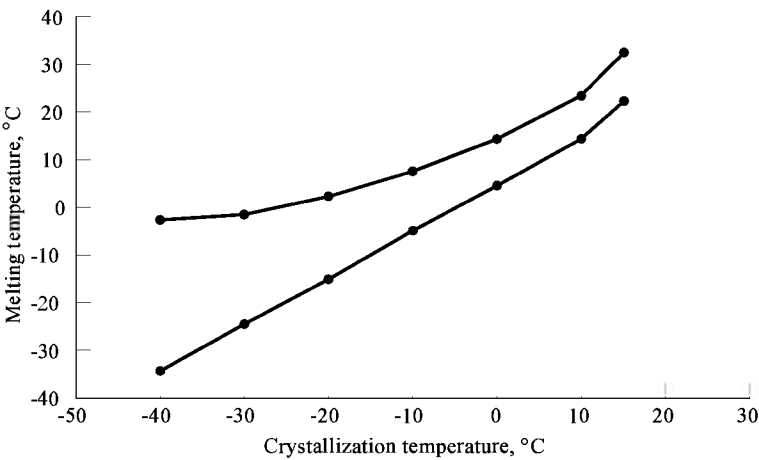


Fig. 4. Relation between crystallization temperature and melting temperature for natural rubber.

Frozen rubber is normally thawed by storing in a hot room at 40–50°C. Because rubber is a relatively poor thermal conductor, melting is governed by the slow conduction of heat through the bale. As the rubber melts, distinct regions of thawed and frozen rubber become apparent and a sharp boundary can be distinguished between the two (18). Palletized frozen rubber can take a considerable time to thaw; in a hot room at 50°C it takes about two weeks for the center of the pallet to reach 30°C. Dismantling frozen pallets and thawing bales singly or in well-spaced stacks obviously helps, but this can be a costly operation. Methods of dealing with frozen rubber vary from factory to factory, and consumer experience determines conditions appropriate to particular needs or limitations.

Storage Hardening. Storage hardening is a slow, irreversible increase in the Mooney viscosity of natural rubber that occurs during storage and transport; an increase of some 30 Mooney units is typical for nonviscosity stabilized grades. Hardening is thought to result from the interaction between the nonisoprenic groups, such as aldehydes, present on the rubber molecule to form cross-links, and as the rubber hardens, the extent of long-chain branching and gel increases. Storage hardening is also accelerated by high temperature and low humidity.

Many mechanisms of storage hardening are proposed and all have been reviewed (19). One of the most favored is the bimolecular aldol reaction between rubber aldehyde groups (5,20). Nevertheless, the involvement of rubber-bound epoxide groups has also been postulated in cross-link formation (21–23), though this was later refuted (24). The hardening mechanism is certainly catalyzed by the presence of amino acids in the latex (25); however, the question of what constitutes a definitive mechanism for storage hardening has yet to be resolved.

Nevertheless, storage hardening can be effectively inhibited by the addition of a reagent such as hydroxylamine, which produces an oxime via a

condensation reaction and thus blocks the cross-linking process (26). For viscosity-stabilized grades (CV) produced in this way, the viscosity increase during transit and storage is negligible.

Cure Characteristics. Methods of natural rubber production and raw material properties vary from factory to factory and area to area. Consequently, the cure characteristics of natural rubber can vary, even within a particular grade. Factors such as maturation, method and pH of coagulation, preservatives, dry rubber content and viscosity-stabilizing agents, eg, hydroxylamine-neutral sulfate, influence the cure characteristics of natural rubber. Therefore the consistency of cure for different grades of rubber is determined from compounds mixed to the ACS1 formulation (27). The ACS1 formulation is as follows: natural rubber, 100; stearic acid, 0.5; zinc oxide, 6.0; sulfur, 3.5; and 2-mercaptobenzothiazole (MBT), 0.5.

This is an activator-starved formulation and so is highly sensitive to the presence of nonrubbers that are capable of activating or accelerating vulcanization, and Table 2 illustrates the cure behavior of different grades of SMR (28). Cup lump grades show the highest state of cure and fastest rate of cure, whereas the stabilized grade, SMR CV, shows the lowest state of cure and slowest cure rate.

Table 2. Cure Characteristics of SMR Grades Mixed to the ACS1 Formulation

Grade	Monsanto rheometer, ^a 160°C, ±3° arc		
	MHR–ML, ^b J/cm ²	T _{S2} ^c , min	T ₉₀ ^d , min
SMR CV	29.4	2.2	11.6
SMR L	33.9	1.8	9.7
SMR 5	37.2	1.5	7.8
SMR 10	40.0	1.3	6.8
SMR 20	41.1	1.2	6.8

^a See RHEOLOGICAL MEASUREMENTS.
^b MHR = maximum stiffness during cure minus minimum stiffness, both in a rheometer. To convert J/cm² to lbf/in., divide by 0.0175.
^c T_{S2} = Scorch time.
^d T₉₀ = Fractional modulus (90% of MHR–ML).

Although filled systems are not as sensitive to variations in this type of cure behavior, certain correlations can be made with practical vulcanizing systems (29).

Plasticity Retention Index. The oxidation behavior of natural rubber may affect both the processing characteristics and final vulcanizate performance, and the plasticity retention index (PRI) test can be used to give an indication of both. Natural antioxidants present in natural rubber give some protection and a measure of the efficacy of protection is given by PRI. PRI% = $\frac{P_{30}}{P_0} \times 100$, where P_0 is the initial Wallace plasticity and P_{30} is the plasticity after aging in an oven at 140°C for 30 min. During mastication, rubber with high values of PRI tends to break down less rapidly than rubber with a low value; normally there is little significant difference between rubbers with a PRI above a value of 60%. PRI can also be correlated with the aging performance of natural rubber vulcanizates (30). For example, raw rubber with a large PRI value gives better high temperature aging resistance, even in a formulation containing two parts of *p*-phenylenediamine antioxidant.

Chemistry and Technology

Natural rubber as obtained from *Hevea brasiliensis* is *cis*-1,4-polyisoprene with small amounts of nonrubber produced by the tree. Although the double-bond structure is useful for chemical modification purposes, it does not benefit natural rubber's heat resistance. Natural rubber is relatively unstable compared to the more modern almost fully saturated elastomers, such as EPDM rubber (see ELASTOMERS, SYNTHETIC ETHYLENE PROPYLENE DIENE RUBBER). In terms of physical properties, the strain crystallizing nature of natural rubber gives it high tensile and tear strengths, and it exhibits both low heat buildup and low rolling resistance. These latter two features make it especially attractive for tire manufacture. However, unless modified, it shows extensive swelling in oils, is relatively permeable to gases, and is not generally suitable for damping applications. In the majority of applications, natural rubber is used in blends and, until fairly recently (31), remarkably little was known about the distribution of cross-links in such blends. Modern techniques have been adapted to look into this subject, with a consequent improvement in properties brought about by achieving a better balance of cross-links in the blend. In latex applications, a large proportion of the industry has moved out to the Far East, for example to Malaysia, if only because it is logical to ship the final product to the end user country rather than the latex, which contains 40% water. In general terms, the proportion of natural rubber in the total rubber market is slowly increasing (ca 1996), partly because it is technically necessary in tires for radial sidewalls, low rolling resistance, and low temperature behavior, and partly because the rubber industry is moving to the Far East, where the bulk of natural rubber is grown and is more easily available.

Vulcanization. Natural rubber is generally vulcanized with a sulfur-based system, the mechanism of which is extremely complex and has been studied for many years by a number of workers (32–35). Basically, a high sulfur, low accelerator (2.5 parts phr of sulfur and 0.5 part of sulfenamide) system results in a mainly polysulfidic cross-linked rubber that has a high tensile strength and exhibits good dynamic performance, but has relatively poor aging characteristics, even with an antidegradant present (Table 3). On the other hand, an efficient vulcanizing (EV) system based on low sulfur and high accelerator levels (see Table 3) leads to a mainly monosulfidic cross-linked network with consequent improvement in compression set and aging characteristics, but less good tensile and tear properties and inferior dynamic performance. Between them there are a multitude of semi-EV systems that fall between the high sulfur–low accelerator and low sulfur–high accelerator recipes and give a compromise in properties, as indicated in Table 3. For maximum heat resistance, as distinct from oxidative aging, there are also the Peroxide and Novor systems. The peroxide vulcanizing agents, eg, dicumyl peroxide [80-43-3], give direct carbon–carbon cross-links for extreme thermal stability. However, such systems have little or no scorch safety, unless a free-radical trap is incorporated, and they tend to require very long cure times. For example, about six times the half-life of the peroxide is needed to ensure complete decomposition; otherwise, aging performance is reduced (36). Further, *p*-phenylenediamine-type antioxidants are to be avoided, as they reduce efficiency of the cure. The virtue of peroxide vulcanization of natural rubber lies in improved heat resistance and good aging, creep, and compression set. In general, however, the vulcanizates suffer from poor low temperature crystallization performance compared to a conventional sulfur cure, and also have inferior tensile and tear properties. Urethane cross-linking systems (37), eg, Novor 950 (see Table 3) are also extremely heat resistant, but exhibit inferior tensile and dynamic properties compared to conventional sulfur-cured vulcanizates. One added virtue is that they can be used in conjunction with sulfur systems to produce an excellent compromise according to the ratios used (38).

Table 3. Properties of Carbon Black-Filled Natural Rubber Vulcanizates With Various Cure Systems^a

Property	Cure system				
	Conven-tion al sulfur ^b	Semi-efficien t sulfur ^c	Efficient sulfur ^d	Peroxide ^e	Novor 950 ^f
other components, phr					
oil	4	4	4	3	
zinc oxide	5	3.5	5	5	5
stearic acid	3	2.5	2		1
hardness, IRHD ^g	65	65	67	61	70

relaxed modulus at 100 ^o o, MPa ^h	2.08	2.22	2.34	2.28	2.60
tensile strength, MPa ^h	28.8	30.1	24.2	21.4	24.0
elongation at break, ^o o	515	485	390	310	460
Dunlop resilience, ¹ ^o o	70	77	67	72	66
ring fatigue, 0–100 ^o o, kHz	223	106	68	51	90
Goodrich HBU, ¹ °C	29	32	36	34	
compression set, 24 h at 70°C, ^o o	27	14	10	11	
retention of tensile strength after 7 d at 100°C, ^o o	27	46	76	51	70

^a 100 phr SMR 5; 50 phr N330 carbon black.

^b Plus sulfur, 2.5 parts; *N*-*t*-butyl-2-benzothiazolesulfenamide, 0.5 part; Santoflex 13, 2 parts.

^c Plus sulfur, 1.2 parts; *N*-cyclohexyl-2-benzothiazolesulfenamide, 0.8 part; tetramethylthiuramdisulfide, 0.4 part; Santoflex 13, 2 parts.

^d Plus sulfur, 0.33 part; 2-morpholinothiobenzothiazole, 3.0 parts; tetramethylthiuramdisulfide, 2.0 parts; Santoflex 13, 2 parts.

^e Plus dicumyl peroxide, 2.5 parts; Flectol H, 2 parts.

^f Plus Novor 950, 6.7 parts; Caloxol, 5 parts; zinc 2-methyldithiocarbamate, 2 parts; zinc 2-mercaptobenzimidazole, 2 parts; 2,2,4-trimethyl-1,2-dihydroquinoline (polymerized), 2 parts.

^g IRHD International Rubber Hardness Degrees (ISO 48).

^h To convert MPa to psi, multiply by 145.

¹ A measure of rebound resilience (BS903: Pt. A8: 1990, Method A).

¹ HBU heat buildup, a measure of temperature rise and resistance to fatigue (ISO 4663-3).

In summary, there are a range of vulcanizing systems which can be used for natural rubber, and the choice is dependent on the combination of properties required. No single one offers ideal, all-around properties combined with good heat resistance. The end user has to be selective, according to the properties required for the final application. Certain properties such as oil resistance and gas permeability have been omitted from Table 3, because in regard to these properties natural rubber is substantially inferior to synthetic rubbers such as acrylonitrile rubber and halobutyl rubber (see ELASTOMERS, SYNTHETIC; RUBBER COMPOUNDING).

Protective Systems. As in the case of sulfur vulcanization, there have been many research workers in the field of natural rubber oxidation and its protection (39,40). The principal problem has always been to find an antioxidant that provides protection against flex-cracking, and yet does not stain. None has been found as of the mid-1990s, hence it is best to categorize the various types according to their capabilities.

The principal category of protective agents is the staining antidegradants, which protect against oxygen, ozone, and flex-cracking (see ANTIOXIDANTS; ANTIOZONANTS; AMINES, AROMATIC; RUBBER CHEMICALS). By far the most effective are the *p*-phenylenediamines. These are the *N*-alkyl-*N'*-phenyl, *N,N'*-dialkyl, and *N,N'*-diaryl compounds. As a general rule, those with the lowest molecular weight are the most reactive, but also the most volatile, most easily extracted, and the worst staining. The second category is staining antioxidants with antiflex cracking capability only. Examples of these are diphenylamine–acetone condensates and naphthylamine derivatives. They are good antiflex cracking agents, but not as effective as the *p*-phenylenediamines. The condensates are good as general-purpose antidegradants and are used in tire compounds. The third type is made up of the low staining amine antioxidants, which include polymerized 2,2,4-trimethyl-1,2-dihydroquinoline, di-β-naphthyl-*p*-phenylenediamine, and alkylated or aralkylated diphenylamines. These impart good protection against heat and oxygen, and the *p*-phenylenediamine is also effective as a metal deactivator. All stain less than most of the other amine antioxidants.

Among nonstaining antioxidants, the least staining and discoloration is given by the phenolic antioxidants, such as the phenol-alkanes, hindered phenols, and hydroquinones. These give reasonably good antioxidant protection, but some are susceptible to pinking on exposure to light. For nonstaining protection against ozone attack under static conditions, waxes (qv) are used. The type of wax is determined by the service condition involved and the temperature. In general, the melting temperature of the wax should not be less than 20°C above the maximum temperature that the surface of the product is likely to reach. The wax must bloom to the surface and form a coherent layer to work satisfactorily. Finally, it is well known that traces of certain metals such as copper, manganese, and iron act as catalysts in the oxidation of natural rubber vulcanizates. Many amines and some phenolics are effective inhibitors of metal-catalyzed oxidation. Here again, *p*-phenylenediamines are particularly effective, providing their staining can be tolerated. Blends of antioxidants are also used and 2-mercaptobenzimidazole or its zinc salt have been found to be very effective.

Modified Forms of Natural Rubber. Because of the limitations in the properties of natural rubber itself, many workers, even before synthetic rubber was produced, attempted to change its structure to broaden its use into new applications. Several of these were only scientific curiosities such as, for example, hydrogenated natural rubber (41,42) and ethyl *N*-phenylcarbamoylazoformate (ENPCAF) modified natural rubber (43). However, these led the way to more practical materials such as chlorinated natural rubber (44,45) for chemical- and heat-resistant paints and coatings, and cyclized natural rubber (46) for shoe solings, hard moldings, and industrial rollers. Although these have virtually disappeared, poly(methyl methacrylate)-grafted natural rubber, superior processing natural rubber, and epoxidized natural rubber are available. The first of these, poly(methyl methacrylate)-grafted rubber (MG rubber), is made by the polymerization of vinyl monomers with the rubber (47) to produce a material of increased modulus or hardness according to the percentage of grafting. The overall result is a modified natural rubber, which is a hard flexible material combining high rigidity with complete recovery from impact (48). MG rubber is used for hard impact-resisting rigid articles and as an adhesive. The latter use is because it contains within its molecular structure both polar poly(methyl methacrylate) and nonpolar polyisoprene components and these make it an attractive material for bonding unlike surfaces, eg, natural rubber with poly(vinyl chloride) (PVC). The second modified form of natural rubber readily available (48,49) is superior processing rubber. This group consists of intimate mixtures of cross-linked and uncross-linked materials prepared by blending vulcanized latex with diluted field latex in proportion to percentage of vulcanized phase required. There are generally five grades available, as shown in Table 4.

Table 4. Grades of Superior Processing Natural Rubber

Grade ^a	Vulcanized phase, ^o o	Raw rubber phase, ^o o	Oil, ^o o
SP20	20	80	0
SP40	40	60	0
SP50	50	50	0
PA57	80	20	40
PA80	80	20	0

^a PA processing aid.

As is implied by their name, superior processing (SP) rubbers provide significant manufacturing advantages. Their two-phase structure increases the stiffness and Mooney viscosity of rubber mixes and at the same time improves flow behavior. The result is faster production rates and greater throughput. Of course, the other advantage of these SP rubbers is that, unlike some other processing aids, these do not affect the final vulcanizate properties, being modified natural rubber themselves. They are used to assist in calendering, extrusion, and reworking and to reduce or eliminate air-trapping problems. They

are particularly advantageous in the liquid curing or salt bath process. Here the normal practice is to extrude the rubber at as high a temperature as possible to ensure rapid vulcanization on entering the salt bath. This inevitably results in a soft extrudate liable to collapse, along with porosity problems. Replacement of half of the rubber with PA80 eliminates these problems, and also reduces die swell (see Table 4).

The third and newest modified natural rubber available is epoxidized natural rubber (ENR). This modification was actually discovered as early as 1922 (50), although the elimination of ring opening and side reactions to give a purely epoxidized material took another 50 years or so to achieve (51). The resulting polymer is a new material, with properties totally different from natural rubber, as indicated in Table 5.

Epoxidized natural rubber is still a strain crystallizing rubber and therefore retains the high tensile strength of natural rubber. However, as can be seen from Table 5, in other respects they have very little in common. The epoxidation renders a much higher damping rubber, a much-improved resistance to oil swelling (insofar as a 50 mol % modified natural rubber has similar oil resistance to a 34% nitrile rubber), and much-reduced air permeability. This latest form of modified natural rubber therefore widens the applications base of the natural material and enables it to seek markets hitherto the sole province of some specialty synthetic rubbers.

Table 5. Physical Properties of Black-Filled Epoxidized Natural Rubber^a

Properties	NR	Epoxidized	
		25% o	50% o
hardness, IRHD ^b	59	56	59
modulus at 300% o, MPa ^c	7.8	6.9	8.8
tensile strength, MPa ^c	27.1	25.9	27.8
elongation at break, % o	550	590	560
compression set, 24 h at 70°C, % o	17	15	17
Dunlop resilience, 23°C, % o	78	60	25
Goodrich HBU, ^d Δ <i>T</i> , °C	44	46	52
volume swelling, 70 h at 70°C, % o			
ASTM No. 1 oil	66	73	5
ASTM No. 2 oil	114	28	6
ASTM No. 3 oil	191	108	21
air permeability at 23°C, m ⁴ (sN) × 10 ¹⁸	27.0	8.0	1.98

^a Filler: N220 black, 30 parts phr.
^b IRHD International Rubber Hardness Degrees.
^c To convert MPa to psi, multiply by 145.
^d HBU heat buildup.

Thermoplastic Natural Rubber. Thermoplastic elastomers are probably the fastest-growing sector of the polymer market in the 1990s because of their easy processibility, efficiency in reuse of scrap or out-of-spec moldings, and general recycling capability. Natural rubber thermoplastic materials are based on blends of natural rubber and polypropylene (52,53). Basically, there are two types: those with a low natural rubber content, which are really only rubber toughened forms of polypropylene, and the softer grades with high natural rubber content, which are truly classed as thermoplastic elastomers. The latter have high strength and good recovery properties, better aging than conventional natural rubber vulcanizates, and excellent ozone resistance. These materials all have to be processed on thermoplastic machinery, and 180°C is regarded as the minimum melt temperature.

The hard (low natural rubber content) grades are best regarded as rubber-modified polypropylenes. Their main attribute is impact resistance, especially at low temperatures, for any given stiffness, as shown in Table 6.

Table 6. Typical Properties of Hard Blends of Thermoplastic Natural Rubber

Property	Blend		
flexural modulus, MPa ^a	700	900	1100
stress at yield, MPa ^a	16	19	24
tensile strength, MPa ^a	30	33	34
elongation at break, % o	650	700	630
Izod impact strength at -30°C, J m ^{b,c}	>640	250	120

^a To convert MPa to psi, multiply by 145.
^b To convert J m to ftlbf in., divide by 53.38.
^c 1-mm notch tip radius.

By employing additives to improve interfacial adhesion and the cohesive strength of the rubber phase, natural rubber can compete with ethylene-propylene rubbers as an impact modifier for polypropylene. These hard grades, containing between 15 and 25% natural rubber, have the potential for use in the automotive and domestic markets, eg, in bumpers, spoilers, grilles, electrical connectors, and floor tiles.

The more important grades of thermoplastic natural rubber, which fall into the olefinic class of thermoplastic elastomers, are prepared with the natural rubber phase partially cross-linked during blending, a process known as dynamic vulcanization. The hardness of the soft blends is controlled by the natural rubber content, and typical properties of those of 50–90 hardness (Shore A) are shown in Table 7.

Table 7. Typical Properties of Soft Blends of Thermoplastic Natural Rubber

Property	Blend				
hardness, Shore A	50	60	69	80	90
modulus at 100% o, MPa ^a	3.1	4.1	5.3	6.3	8.2
tensile strength, MPa ^a	6.5	8.8	11.2	13.2	15.3
elongation at break, % o	285	315	340	370	505
tear strength, Die C, N mm ^b	22	31	38	43	51
tension set, % o	9	10	14	18	23
compression set, 24 h at 70°C, % o	30	35	36	41	55

retention of tensile strength after 7 days at 100°C, %	96–98	96–98	96–98	96–98	96–98
--	-------	-------	-------	-------	-------

^a To convert MPa to psi, multiply by 145.
^b To convert N/mm to ppi, divide by 0.175.

The good strength properties of the natural rubber thermoplastic blends may relate to the natural rubber phase undergoing strain-induced crystallization, as there is a distinct upturn in the stress–strain curve of the natural rubber blend which is not seen with other olefinic thermoplastic elastomers. It may be noted too that the natural rubber blend exhibits good recovery properties with compression set values of the softer materials not too dissimilar to that of a conventional high sulfur vulcanizate (see Table 3). Thermoplastic natural rubber, when correctly compounded, has excellent aging characteristics, as shown in Table 7, and ozone resistance is also high. Even the softest grades with the highest natural rubber content exhibit no cracking after exposure to 100 pphm ozone for 7 d at 40°C and 20% strain. Its recyclability is demonstrated in Table 8 where the retention of tensile strength and modulus is shown after reprocessing five times.

Table 8. Effect of Reprocessing Thermoplastic Natural Rubber

Property	Percentage retention				
	1 Recycle	2 Recycle	3 Recycle	4 Recycle	5 Recycle
modulus at 100% extension	94	92	90	87	84
tensile strength	105	103	103	99	100
elongation at break	104	107	109	108	112

These thermoplastic natural rubber elastomers have a place in the modern world, where recycling has become so important, and when excessive heat is not found in service. Thus, footwear, glazing seals, sports goods, hose, domestic products, and a whole range of automotive products have already been identified for such use. It must be noted, however, that tires are not a potential market for these materials, because of the high temperatures which result from emergency braking.

Natural Rubber–Rubber Blends. The majority of natural rubber is used in blends with other elastomers. This has been common practice for many years, and most of the formulations employed by the manufacturers of rubber products have been derived empirically. It was not until 1988 that there was a method of measuring cross-link density in blends (31,54). Using a swollen-state nmr spectroscopy technique, it has been found possible to not only measure cross-link density in natural rubber itself, but also in the individual phases of vulcanized blends. This technique has enabled studies to be made of a range of natural rubber–synthetic rubber blends and, as was expected to be the case, has also shown a total imbalance of cross-link density in incompatible blends such as natural rubber and nitrile rubber (55). However, the ability to measure individual cross-link densities in such a blend and effectively confirm that the natural rubber phase is under-cross-linked enables changes to be made to redress the imbalance. Thus, by careful selection of suitable curatives, which are likely to be more soluble in the natural rubber phase, a much more even distribution of cross-links between the two polymers has been achieved, with the resulting improvement in physical properties such as tensile strength. The technique is being applied to other natural rubber–synthetic rubber blends, and it seems likely that improvements will even be made in blends of natural rubber with EPDM. This will be of immense benefit, because a combination of the virtues of these two polymers, the strength and resilience of natural rubber and the oxidation resistance of EPDM, will provide a material much sought-after by the manufacturing industry.

Applications

In the same way that natural rubber is predominantly used in blends, it is also predominantly used in tire manufacture. Its excellent building tack, low heat buildup, low rolling resistance, and good low temperature performance make it the polymer of choice for many parts of tire construction, for both passenger and truck vehicles. The effects of radialization and demand for low rolling resistance and good low temperature performance have all tended to benefit natural rubber, especially in truck tire construction, as shown in Table 9.

Table 9. Average Polymer Composition of Truck Tire Components,^a %

Year	Tread			Sidewall			Ply belt		
	NR	SBR	BR	NR	SBR	BR	NR	SBR	BR
1974	45	21	34	48	37	15	71	20	9
1981	60	12	28	44	19	37	84	11	4
1983	77	7	16	58	6	36	100		
1985	86	5	9	62		38	100		
1990	86	5	9	75		25	100		
1994	100			60		40	100		

^a NR = natural rubber, SBR = styrene–butadiene rubber, and BR = butadiene rubber.

A large proportion of the circa 3.7 × 10⁶ t/yr of natural rubber used in tires is consumed in truck tires, off-the-road tires, and aircraft tires, all of which demand a low heat buildup performance. The retreading of truck tires was also the province of natural rubber until the precured process was developed. Prior to the advent of this technique, the tack of natural rubber was essential for the unvulcanized rubber to adhere onto the buffed carcass, and the thickness of the shoulder region was such that the low heat buildup of natural rubber was essential to ensure that no failure occurred in service. With the precured tread process neither of these attributes were necessary and, synthetic rubber, which was and continues to be used, was found to give particularly good wear performance under low severity conditions, especially in the United States. However, work (56) has shown that natural rubber-based formulations can be developed which give a similar order of wear performance to the all-synthetic rubber tread, but with the additional benefit of lower rolling resistance, and hence better fuel economy.

Natural rubber was also used extensively in its oil-extended form in winter tires in the 1970s (57). Use of oil-extended natural rubber treads, found to have excellent traction on ice and snow, superseded studded synthetic rubber treads when studs were banned in certain countries and states owing to the damage they cause to partially cleared roads. This concept has been extended into all-season tires, which account for over 75% of original equipment and replacement tires in the United States. It has been shown (58) that part replacement of styrene–butadiene rubber (SBR) in the formulation of all-season tire tread compounds with oil-extended natural rubber increases ice and snow traction, reduces rolling resistance, and has no effect on normal wet grip. Also, there is only a minor trade-off in wear performance, because below a tire surface temperature of approximately 32°C, the wear of natural rubber is superior to SBR, whereas above this temperature the reverse is true (59). Thus, wear of an all-season tire ultimately depends on the surface temperature of the tread over its annual cycle of temperatures.

Other than tire use, there are few other significantly large application areas for natural rubber that can be identified, as indicated in Table 10. The use of natural rubber in latex products covers items such as gloves, condoms, balloons, catheters and other dipped goods, latex thread, foam and carpet backing, and rubberized coir and hair. In total, latex goods consume about 11% of world rubber production. Thereafter, there are four categories: footwear,

nonautomotive engineering, belting and hose, and nontire automotive applications, which consume 1.5–4° º each. Footwear is self-explanatory, with the market for athletic shoes, especially those made in the Far East, tending to keep natural rubber in this market. Nonautomotive engineering applications cover the use of natural rubber in civil engineering products such as bridge bearings, off-shore oil field equipment, and the growing market for rubber–metal-laminated bearings to protect buildings against earthquakes (60,61). This relatively new application was born out of the early work on metal-laminated natural rubber bearings for bridges, which later spawned the use of natural rubber bearings to protect buildings against ground-borne vibrations, eg, from the London Underground. This in turn developed in the 1980s into natural rubber bearings to protect buildings against earthquakes, a technique fast developing around the world in earthquake-prone areas, notably Japan, New Zealand, Italy, and the state of California in the United States. Another significant use of natural rubber is in belting and hose, where natural rubber's high strength, resilience, and fatigue performance still help to maintain its use as a premium rubber for these applications.

Table 10. End Uses of Natural Rubber

Use	Amount, 10 ³ t	Percent of total
tires and related products	3800	71
latex products	585	11
footwear	210	4
nonautomotive engineering	200	4
belting and hose	80	1.5
automotive (nontire)	80	1.5
wire and cable	15	<1
other	360	
Total world natural rubber consumption	5333	100

The nontire automotive sector includes engine mounts, where natural rubber is still regarded as the preferred polymer, and the many other rubber bearings, seals, grommets, washers, and boots used in the automotive industry for which oil-resistance is not required. It is likely that some of these items will change to thermoplastic materials, including thermoplastic natural rubber, in coming years as the demand for total recyclability of automobiles takes hold (see RECYCLING). The wire and cable industry has continued to use less natural rubber over the years because of its poor durability as compared to PVC and ethylene–propylene rubbers. However, improvements in blend technology may well reverse this trend if natural rubber ethylene–propylene rubber blend properties can be optimized for the cable market.

The other miscellaneous uses of natural rubber cover a multitude of applications, none of which of themselves consume a large tonnage. Many of these are traditional uses such as pipe seals, medical closures, rollers, small solid tires, mountings for a whole range of domestic and commercial appliances, rubber balls and tubing, milking inflations, and other agricultural-based applications. In total these make up approximately another 7° º of natural rubber usage, which for 1993 had a total market share of some 5.33×10^6 t.

Natural Rubber Latex

Preparation. Natural rubber latex has a pH of 6.5–7.0 when it emerges from the tree. Its rubber content is between 25–40° º depending on the clone soil conditions, weather conditions, and time of tapping. From the moment the latex comes in contact with air, bacteria and enzymes act to reduce the pH and destabilize it. If left unchecked, the bacterial action causes the latex to coagulate within hours. A preservative must therefore be added either to the cup prior to tapping or during collection soon after flow has ceased. The traditional, and still widely used, field preservative is ammonia, which is added to the latex as a small quantity of a fairly concentrated solution. The drawback of ammonia is that it evaporates from aqueous solution, which can be unpleasant for users and can lead to changes in concentration during storage and transport. Ammonia is only effective as a bactericide above a concentration of about 0.2° º; below 0.1° º concentration it can actually promote bacterial growth (62). For these reasons, a 1:1 mixture of tetramethylthiuram disulfide and zinc oxide (TZ) has been used increasingly as a field preservative.

At the processing factory, the field latex is bulked into batches of at least several tons and ammoniated to a level of about 0.5° º, usually with ammonia gas. Bulking has the effect of ironing out variations in rubber content and other properties of latex from various sources. Ammonia has the role of preservative, but also raises the pH of the latex which promotes hydrolysis of some natural lipids to release stabilizing fatty acid soaps. Diammonium phosphate can be added at this stage to precipitate any magnesium ions in the latex as magnesium ammonium phosphate, which settles as a sludge (61).

Concentration. The most common method of concentrating field latex is centrifugation. The two widely used types of centrifuge are the de Laval and Westfalia. Both operate by spinning the latex at high speed between a set of closely spaced conical disks. Because rubber is less dense than the water-based serum, the rubber particles tend to be concentrated toward the center of the centrifuge and the serum toward the outside. In a continuously running machine, the rubber content of the concentrate depends on the rubber content of the incoming latex and the feed rate. The output from the low rubber side of the centrifuge is called skim and generally contains about 5° º rubber. This is coagulated and processed to produce skim rubber, which is considered a low grade of dry rubber owing to its high ratio of nonrubbers to rubber. After centrifuging, the rubber content of the concentrate is adjusted to 60° º and final preservation is carried out. This consists of either raising the ammonia concentration to 0.7° º to produce high ammonia (HA) or full ammonia latex, or keeping the ammonia level low (LA) to about 0.2° º and adding a secondary preservative. The most common secondary preservative is TZ (LATZ latex) (63). The use of others, such as boric acid (LA–BA latex) and sodium pentachlorophenate (LA–SPP), has declined owing to concerns about toxicity of these products.

Although more than 90° º of all latex concentrate is centrifuged, small quantities are produced by creaming and evaporation. Creaming takes advantage of the natural tendency of the rubber particles to rise to the surface of static latex. Commercially, creaming agents such as ammonium alginate are added to increase the rate of creaming (64). The process is of necessity a batch process, and it can take several weeks to achieve the required concentration.

Evaporated latex is produced in continuous evaporators, which use a combination of heat and reduced pressure to remove water from the latex. The main distinguishing features of evaporated latex is that 100° º rubber recovery is achieved, but unlike the other methods, the nonrubbers are also concentrated. The most well-known evaporated latex is Standard Revertex (trade name from Revertex).

Latex Properties. Almost all latex concentrates are produced to meet an international standard, ISO 2004 (Table 11). This standard gives minimum requirements for rubber content and sets limits on some other latex properties. One of the most important properties is mechanical stability, which gives an indication of the resistance to flocculation or coagulation during processing of the latex. Mechanical stability rises naturally, particularly during the first month after concentration, owing to the chemical action of ammonia on some latex compounds (65). It can also be raised artificially by the addition of soaps to the latex. Copper and manganese concentrations are limited because of the severe degrading effects of these metals on finished products. The volatile fatty acid (VFA) number gives an indication of bacterial activity and thus shows how well the latex has been preserved (66). This is a particularly important test for low ammonia latices, which have no minimum requirement for ammonia content. For all five types in Table 11, the VFA specification (ISO 506) is as agreed by the interested parties, but not to exceed 0.20. The KOH number (ISO 127) for all five types is as agreed by the interested parties, but not to exceed 1.0. If the latex contains boric acid, the KOH number may exceed the specified value by an amount equivalent to the boric acid content as determined according to the method specified in ISO 1802. Color and odor specification are as follows: on visual inspection there should be no pronounced blue or gray; and after neutralization with boric acid, there should be no pronounced odor of putrefaction. The most important latex properties not covered by ISO 2004 are latex viscosity, magnesium content, and stability in the presence of zinc ions (67,68).

Processing and Products. The main distinguishing feature of rubber products made from latex rather than dry rubber is the rubber thickness, which is limited to a few millimeters. The diffusion of water through rubber is so slow that the drying times of thicker-walled products would be impractically long.

In producing latex products, the chemicals required for vulcanization, stiffening, coloring, antioxidant protection, or other purposes are added as solutions, emulsions, or fine dispersions to the latex before forming the product. Because no heat is generated during this mixing, it is possible to use ultrafast accelerators that would cause scorch problems in dry rubber compounds. Indeed, it is desirable to use ultra-accelerators in latex products because it allows lower vulcanization temperatures to be used, which reduces the potential of oxidative degradation in these thin-walled products. The most widely used accelerators for latex products are the dithiocarbamates, particularly zinc diethyldithiocarbamate (ZDEC) and zinc dibutyl dithiocarbamate (ZDBC or ZBUD). Thiazoles such as zinc mercaptobenzothiazole (ZMBT) are often used as secondary accelerators with dithiocarbamates, and thiurams such as TMTD are sometimes used in place of dithiocarbamates, especially where low sulfur systems are required. Another effect of wet mixing of chemicals that distinguishes latex products from dry rubber products is that mastication does not occur and hence the original high molecular weight of the polymer is retained.

Table 11. ISO Specifications for Natural Rubber Latex Concentrates

Characteristic ^a	Limits					Method of test
	Type HA	Type LA	Type XA	Type HA, creamed	Type LA, creamed	
total solids content ^b , %	61.5	61.5	61.5	66.0	66.0	ISO 124
dry rubber content ^b , %	60.0	60.0	60.0	64.0	64.0	ISO 126
nonrubber solids ^c , %	2.0	2.0	2.0	2.0	2.0	
alkalinity (as NH ₃), % on latex concentrate	0.60 min	0.29	0.30 min	0.55 min	0.35	ISO 125
mechanical stability ^d , s	650	650	650	650	650	ISO 35
coagulum content, %	0.05	0.05	0.05	0.05	0.05	ISO 706
copper content, mg kg of total solids	8	8	8	8	8	ISO 8053
manganese content, mg kg of total solids	8	8	8	8	8	ISO 7780
sludge content, %	0.10	0.10	0.10	0.10	0.10	ISO 2005

^a Value is maximum unless otherwise specified.
^b The requirement is for either total solids content or dry rubber content. Value is minimum.
^c The difference between the total solids content and dry rubber content.
^d A minimum mechanical stability may be required which is greater than the minimum value specified.

The most important group of latex products are the dipped goods, which account for more than 50% of all latex concentrate usage. Dipped goods are produced by dipping a shaped former into a suitably formulated latex compound, and then withdrawing it. The latex deposit is dried and vulcanized in hot air to give the product, which is then stripped from the former. There are three main variations on the dipping process, ie, straight, coagulant, and heat-sensitive dipping. Straight dipping gives the thinnest deposits and is used to produce condoms. The thickness is dependent primarily on the latex viscosity and rubber content. In coagulant dipping, the former is dipped in an inorganic salt solution prior to the latex dip. Thicker deposits can be produced this way and the final thickness is dependent on the latex dwell time, as well as other factors. Medical, household, and industrial gloves are produced by this method, as are many other products, such as balloons, bladders, catheters, and other medical devices. Heat-sensitive dipping is used to make the thickest latex products, such as baby-bottle nipples. It requires hot formers to be dipped into a suitably compounded heat-sensitive latex.

Aside from dipping, the other main products produced from natural rubber latex are elastic thread and foam products. Latex thread is produced by extruding a suitable latex compound under gravity through a tube of the appropriate diameter into an acid bath. On coming into contact with the acid, the latex gels. It is then taken through washing baths and a drying vulcanization oven in a continuous process that results in reels of thread. This thread is used primarily in the garment industry. Latex foam is produced by mechanically incorporating air into a stabilized latex compound until the required density is achieved. The foam is gelled by either a delayed action chemical or by a heat gelation process before vulcanization and drying. The main applications of latex foam are in molded products such as mattresses and cushions, and spread foam for carpet backings. Unlike the dipped products industry, the latex foam industry has suffered from competition from synthetic styrene–butadiene rubber (SBR) latex. Many foam products can be made from either NR, SBR, or blends of the two.

Natural rubber latex is also used in adhesives for tape, packaging, envelopes, and in the footwear industry. It is used in the carpet industry as a binder for backing compounds, but this is another area in which synthetic SBR latex has competed effectively. There are a number of relatively small and specialized applications for natural rubber latex including rubberized coir or hair and cast products such as toys. Latex sheeting which is used in dental dams and for numerous other purposes can be made by dipping or casting onto a continuous belt.

Latex Allergies. Since the late 1980s there has been a great deal of attention given to the problems of allergic reactions to latex products. One reason for this is that a sharp increase in consumption of latex products in body contact applications occurred, notably medical gloves and condoms, largely as a result of concern about the spread of AIDS. In addition, there may well have been an increased incidence of sensitization to the latex allergy as a result of some poor-quality products being produced to meet the sudden upturn in demand. These factors, combined with a general increase in susceptibility of Western hemisphere populations to allergies, and a spread in awareness caused by a few publicized cases, caused an apparent eruption in this problem despite the fact that latex products of essentially the same composition had been around for decades previously.

There are two types of allergy that can be caused by latex products. Type I allergy or immediate hypersensitivity reactions are mediated by Immunoglobulin E (IgE), which is produced in the body of susceptible individuals in respect to contact with the allergen. Individuals carrying latex-specific IgE are considered to be sensitized to the allergy, and subsequent contact with latex products is likely to induce an immediate allergic response (69). The clinical manifestations of the reactions range from urticaria (nettle rash), which is the most common symptom, to anaphylaxis, which is rare but can result in loss of consciousness (70). It is clear that the components of latex products which elicit the Type I allergy are proteins and protein residues which remain in the product after manufacture. These substances do not contribute in any significant way to the properties or performance of the product but are a result of natural origin of the material. Latex product manufacturers have therefore put considerable effort into finding ways to remove or reduce the levels of allergenic proteins (71). Extensive washing or leaching, chlorination, and polymer coating of the surface are some of the main ways in which this has been achieved.

Because of the importance of residual proteins in causing Type I allergies, methods for their analysis have also been an area of great interest. Extractable proteins in latex products can be determined either chemically or by immunological methods. The chemical methods, such as the Bradford and Lowry assays, do not distinguish between allergenic and nonallergenic proteins, whereas the immunochemical methods are specific to either allergenic or antigenic proteins. Of the chemical methods of analysis, a modified Lowry method (72) has gained widest acceptance and one version of this method has been adopted by ASTM (73). Immunological methods of analysis such as radioimmunosorbant assays (rast) and enzyme-linked immunosorbant assays (elisa) are more specific and much more sensitive than the chemical methods, but require a large, homogeneous, and representative pool of sera from sensitized individuals in order to give consistent results.

The other type of allergy that can be attributed to natural rubber latex products is a Type IV allergy. This allergy also requires sensitization of the

individual but the allergic response when a sensitized person comes into contact with the allergen is delayed by hours or days. The reactions are limited to localized skin complaints but the reaction can be sufficiently uncomfortable to prevent a sufferer from using natural latex products. The causes of the Type IV allergy are accelerators and antioxidants, particularly the dithiocarbamate accelerators, which are added to the latex during product manufacture. Products containing very low levels of residual accelerators have been described as hypoallergenic; however, the use of this term is discouraged, because it does not distinguish between the two types of allergy.

Economic Aspects

Malaysia, Indonesia, and Thailand are the three main rubber-producing countries in the world; in 1994 they contributed 73% to the world's natural rubber production, which was some 5.7×10^6 t (74). In the early 1990s the relative consumption of natural rubber has remained fairly constant at 38–39% of total rubber consumption, despite competition from synthetic rubber.

Figure 5 shows natural rubber production from 1980 to 1994 split between the three principal producing countries, Malaysia, Indonesia and Thailand, and "others." The others constitute production from (in alphabetical order) Brazil, Cambodia, Cameroon, China, Côte d'Ivoire, Gabon, India, Mexico, Nigeria, the Philippines, Sri Lanka, and Vietnam, all of which produce a varied amount of natural rubber. Figure 5 shows that Malaysian production peaked in 1988 at 1.66×10^6 t and is down to just over 1×10^6 t yr. This has been brought about by many years of low prices causing smallholders to abandon their trees and seek employment in the fast-developing downstream industries in Malaysia, combined with a general labor shortage in the country. This has allowed Thailand to become the No. 1 natural rubber producer, with an output of 1.72×10^6 t in 1994, compared to Indonesia at 1.36×10^6 t, and Malaysia at 1.10×10^6 t. Thailand is also industrializing and it is likely to reach a peak in production in the latter half of the 1990s. Indonesia still has considerable land available for rubber planting and a plentiful labor supply, so production there is likely to increase. Nevertheless, it is difficult to see where the extra natural rubber required for the year 2000 will come from, with a predicted consumption by then of some 6×10^6 t. However, world demand for natural rubber will no doubt control its price, which in turn will dictate the volume produced. Much will depend on the world economy and whether there continue to be further International Rubber Agreements prior to the year 2000.

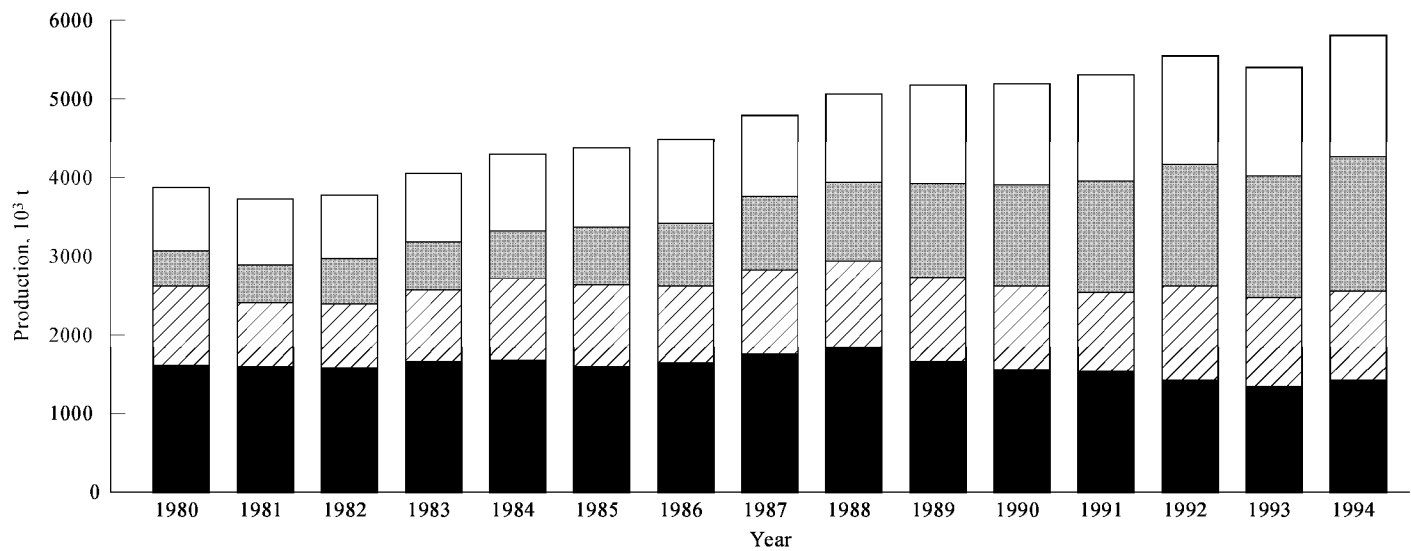


Fig. 5. Natural rubber production, where (■) represents Malaysia; (▨) Indonesia; (▩) Thailand; and (□) other.

The natural rubber industry is important in the producing countries, providing work and earning money from exports. Natural rubber and rubber products earned 4.1% of the export income for Malaysia in 1994, of which 2.1% was derived from natural rubber. A total of 1.7×10^6 ha of land were under rubber but only 1.1×10^6 ha yielded rubber, owing in part to a labor shortage. In Malaysia, about 84% of the total area planted with rubber was tapped by smallholders, who produced 72% of the rubber, with the rest produced by estates (75,76). Indonesia (77) has a similar proportion of land (83%) cultivated by smallholders. Smallholders are becoming increasingly important for rubber production, and account for 73% of the total world area under rubber cultivation with Thailand almost 96%, and Sri Lanka, 56%. Rubber smallholders are typically self-employed farmers who depend on their land to support their families and also plant other crops to supplement their income. Estates are self-contained units owned by individuals or large companies, and large estates carry out all stages of rubber production. Smaller estates may send their rubber to other factories for processing before export.

Consumption of NR Latex. The total world consumption of natural rubber latex was 585,000 t in 1993, more than double that of 10 years earlier. The proportion of total natural rubber used as latex concentrate also increased from 7% in 1983 to 11% in 1993. Malaysia remains the dominant latex concentrate producing country in the world (Table 12). Its decline in exports during the 1980s and early 1990s has been partly compensated by an increase in consumption within the country (Table 13). Since the 1970s there has been a gradual shift in latex consumption from the traditional consumers in Europe and the United States to Asian countries. In 1994 Malaysia consumed 175,000 t of latex, more than the combined consumption of Western Europe and the United States (see Table 13).

Table 12. Net Exports of NR Latex, 10³ t

Country	1989	1990	1991	1992	1993	1994
Malaysia	241	192	190	152	120	119
Liberia	106	40	19	28	3	1
Thailand	26	46	61	69	137	132
Indonesia	34	32	59	39	40	35

Table 13. Consumption of NR Latex, 10³ t

Area	1989	1990	1991	1992	1993	1994
Malaysia	83	112	125	149	151	175
United States	77	74	65	82	73	73
Western Europe	94	78	71	68	67	71
China	42	31	36	54	49	80
India	36	39	35	51	52	54

Korea	29	29	26	25	24	26
Taiwan	25	24	36	32	23	19

BIBLIOGRAPHY

"Rubber, Natural" in *ECT* 1st ed., Vol. 11, pp. 810–826, by T. H. Rogers, Jr., The Goodyear Tire & Rubber Co., Inc.; in *ECT* 2nd ed., Vol. 17, pp. 660–684, by T. H. Rogers, The Goodyear Tire & Rubber Co., Inc.; in *ECT* 3rd ed., Vol. 20, pp. 468–491, by D. R. St. Cyr, Goodyear Tire & Rubber Co., Inc.

1. P. R. Wychedey, *The Planter* **44**, 127 (1968).
2. H. N. Ridley, *Agric. Bull. Malay. Pen.* **7**, 136 (1897).
3. P. D. Abraham and R. A. Tayler, *Trop. Agriculture*, **44**, 1 (1967).
4. A. Aziz bin S. A. Kadir, *Rubb. Chem. Technol.* **67**, 537 (1994).
5. A. Subramaniam, in A. Aziz bin S. A. Kodir, ed., *Proceedings of International Rubber Technology Conference, Kuala Lumpur, 1993*, Rubber Research Institute of Malaysia (RRIM), p. 19.
6. A. Subramaniam, *Technology Bulletin No. 4*, Rubber Research Institute of Malaysia, 1980.
7. J. B. Gomez, *Physiology of Latex (Rubber) Production*, MRRDB Monograph No. 8., Malaysian Rubber Research & Development Board, Kuala Lumpur, Malaysia, 1983.
8. *RRIM Training Manual on Tapping, Tapping Systems and Yield Stimulation of Hevea*, RRIM, Kuala Lumpur, Malaysia, 1980.
9. P. D. Abraham and co-workers, *J. Rubb. Res. Int. Malaya*, Pts. I, II, and III **23**, 85, 90 and 114 (1971).
10. "The Green Book," *International Standards of Quality and Packing for Natural Rubber Grades*, Rubber Manufacturers Association, Inc., Washington, D.C., 1975.
11. R. Tharmalingam, M. C. S. Perera, and R. Gunathilake, in *Proceedings of IRCIND Conference*, India, 1980.
12. M. C. S. Perera, T. A. S. Siriwardene, and H. K. D. C. S. Jayawardene, *Plast. Rubb. Process. Applic.* **12**, 153 (1989).
13. *Revisions to Standard Malaysian Rubber Scheme*, *SMR Bull.* **11** (1991).
14. Indonesian Ministry of Trade Decree No. 184 Kp VI 88, Jakarta, 1988.
15. ISO 2000, *Rubber, Raw Natural—Specification*, International Organization for Standardization, Geneva, Switzerland, 1989.
16. ISO 4660, *Rubber, Raw Natural—Color Index Test*, International Organization for Standardization, Geneva, Switzerland, 1991.
17. E. H. Andrews and A. N. Gent, in L. Bateman, ed., *The Chemistry and Physics of Rubber-like Substances*, Maclaren & Sons Ltd., London, 1963, p. 225.
18. G. M. Bristow and A. G. Sears, *NR Technol.* **13**, 73 (1982).
19. D. R. Burfield, *J. Nat. Rubb. Res.* **1**, 202 (1986).
20. D. R. Burfield and T. C. Tan, *Polym. Comm.* **28**, 190 (1987).
21. D. R. Burfield, *Nature* **249**, 29 (1974).
22. D. R. Burfield and S. N. Gan, *J. Polymer Sci. Polym. Chem. Ed.* **13**, 3725 (1975).
23. D. R. Burfield and S. N. Gan, *J. Polymer Sci. Polym. Chem. Ed.* **15**, 2721 (1977).
24. A. Subramaniam and W. S. Wang, *J. Nat. Rubb. Res.* **1**, 58 (1986).
25. M. J. Gregory and A. S. Tan, in *Proceedings of International Rubber Conference*, Kuala Lumpur, Malaysia, Vol. IV, 1975, p. 28.
26. B. C. Sekhar, *J. Polym. Sci.* **48**, 133 (1960).
27. ISO 1658: *Natural Rubber - Test Recipes and Evaluation of Vulcanization Characteristics*, International Organization for Standardization, Geneva, Switzerland, 1973.
28. *Rubb. Devel.* **39**, 23 (1986).
29. G. M. Bristow, *NR Technol.* **11**, 59 (1980).
30. G. M. Bristow, *Rubb. Devel.* **43**, 24 (1990).
31. P. S. Brown, M. J. R. Loadman, and A. J. Tinker, *Rubb. Chem. Technol.* **65**, 744 (1992).
32. D. K. St. Cyr, in J. I. Kroschwitz, ed., *Encyclopedia of Polymer Science and Engineering*, Vol. 14, Wiley-Interscience, New York, 1988, pp. 687–716; M. Porter and co-workers, in N. M. Bikales, ed., *Encyclopedia of Polymer Science and Technology*, Vol. 12, John Wiley & Sons, Inc., New York, pp. 195–210.
33. A. V. Chapman and M. Porter, in Roberts, ed., *Natural Rubber Science and Technology*, Oxford University Press, Oxford, U.K., 1988, pp. 511–620.
34. A. Y. Coran, in J. I. Kroschwitz, ed., *Encyclopedia of Polymer Science and Engineering*, Vol. 17, Wiley-Interscience, New York, 1989, pp. 666–698.
35. M. R. Krejsa and J. L. Koenig, *Rubb. Chem. Technol.* **66**, 376 (1993).
36. C. S. L. Baker, in *Symposium Proceedings Session B: Materials and Compounding*, Swedish Institution of Rubber Technology Fall Meeting, Boras, Sweden, 1979.
37. C. S. L. Baker, *Kaut. Gummi Kunstst.* **36**, 677 (1983).
38. C. S. L. Baker, *Vulcanization with Urethane Reagents*, NR Technical Bulletin, Malaysian Rubber Producers' Research Association, Brickendonberry, U.K., 1978.
39. R. W. Keller, *Rubb. Chem. Technol.* **58**, 637 (1985).
40. D. Barnard and P. M. Lewis, in Ref. 33, p. 621.
41. H. Staudinger and W. Feisst, *Helv. Chim. Acta* **13**, 1361 (1930).
42. D. R. Burfield, K. L. Lim, P. K. Seow, and C. T. Loo, in J. C. Rajarao and L. L. Amin, eds., *Proceedings of International Rubber Conference*, Kuala Lumpur, Malaysia, Vol. 2, 1985, p. 47.
43. D. Barnard, K. Dawes, and P. G. Mente, *Proceedings of International Rubber Conference*, Kuala Lumpur, Malaysia, Vol. 4, 1975, p. 215.
44. D. E. Woods, *J. Soc. Chem. Ind.* **68**, 343 (1949).
45. G. F. Bloomfield, *The Applied Science of Rubber*, Edward Arnold Ltd., London, 1961, pp. 112–121.
46. G. F. Bloomfield and S. G. Stokes, *Trans. Inst. Rubber Ind.* **32**, 172 (1956).
47. P. W. Allen, in Ref. 17, p. 97.
48. *High Performance Natural Polymers*, Leaflet by Asiatic Development Berhad, Wisma Genting, Jalan Sultan Ismail, Kuala Lumpur, 1993.
49. Rubber Research Institute of Malaya, *Planters Bull.* **32**, 83 (1957).
50. R. Pummerer and P. A. Burkard, *Ber.* **55**, 3458 (1992).
51. I. R. Gelling, *J. Nat. Rubb. Res.* **6**, 184 (1991).
52. I. R. Gelling and A. J. Tinker, in L. L. Amin and L. K. Thong, eds., *Proceedings of the International Rubber Technology Conference*, Penang, Malaysia, 1988, p. 212.
53. D. J. Elliott, *Thermoplastic Elastomers from Rubber–Plastic Blends*, Ellis Horwood Ltd., Chichester, U.K., 1990, pp. 102–129.
54. A. J. Tinker and M. J. R. Loaman, *Rubb. Chem. Technol.* **62**, 234 (1989).
55. P. S. Brown, S. Cook, S. A. Groves, M. V. Lewan, and A. J. Tinker, "Natural Rubber Blends," paper presented to the *ACS Rubber Division Meeting*, Orlando, Fla., Oct. 1993.
56. "Development of Natural Rubber-Based Truck Tyre Retreading Compounds," in M. E. Cain and Siti Zubaidah bte Mohd Rashid, eds., *Proceedings of a UNIDO-sponsored Workshop, Kuala Lumpur, Malaysia*, Malaysian Rubber Producers' Research Association, Hertford, U.K., 1992.

57. J. McLellan, *Rubber Devel.* **24**, 77 (1971).
58. R. Newell and I. R. Wallace, *Kaut. Gummi Kunstst.* **45**, 380 (1992).
59. K. A. Grosch, *J. Inst. Rubber Ind.* **1**, 35 (1967).
60. V. A. Coveney, C. J. Derham, K. N. G. Fuller, and A. G. Thomas, in Ref. 33, pp. 938–957.
61. V. A. Coveney, K. N. G. Fuller, and A. G. Thomas, *Int. Const.*, **31**, 21 (1992).
62. W. L. Resing, in T. H. Messenger, ed., *Proceedings of the Rubber Technology Conference, London*, W. Heffer & Sons, Ltd., Cambridge, U.K., 1954, p. 50.
63. C. J. John, M. Nadarajah, P. S. Rama Rao, C. M. Lau, and C. S. Ng, *Proceedings of the International Rubber Conference*, Vol. 4, Kuala Lumpur, Malaysia, 1975, p. 339.
64. I. H. Duckworth, in *Proceedings of Planters Conference*, Vol. 12, Kuala Lumpur, Malaysia, 1964, p. 134.
65. H. M. Collier, *Trans. IRI* **31**, 166 (1955).
66. D. H. Taysum, *Proceedings of Natural Rubber Conference, Kuala Lumpur*, Vol. 1, Rubber Research Institute of Malaysia, 1960, p. 834.
67. A. D. T. Gorton and T. D. Pendle, in J. C. Rajarao and L. L. Amin, eds., *Proceedings of the International Rubber Conference*, Kuala Lumpur, Malaysia, Vol. 2, 1985, p. 468.
68. A. D. T. Gorton and T. D. Pendle, *J. Nat. Rubber Res.* **1**, 122 (1986).
69. K. Turjanmaa, K. Laurila, S. Makinen-Kiljimen, and T. Reunala, *Contact Dermatitis* **19**, 362 (1988).
70. D. A. Levy, D. Charpin, C. Pecquet, F. Leynadier, and D. Vervloet, *Allergy* **47**, 579 (1992).
71. S. J. Dalrymple and B. G. Audley, *Rubb. Devel.* **45**, 51 (1992).
72. Y. Faridah and H. Y. Yeang, *J. Nat. Rubber Res.* **7**, 206 (1992).
73. *ASTM 5712-95*, American Society for Testing and Materials, Philadelphia, Pa., 1995.
74. *International Rubber Study Group (IRSG) Statistical Yearbook*, London, 1995, p. 2.
75. *Economic Report*, Ministry of Finance, Kuala Lumpur, Malaysia, 1994–95.
76. *Statistics on Commodities*, Ministry of Primary Industries, Malaysia, July 1995.
77. The Economist Intelligence Unit, *Rubber Trends*, 1994, 1st Qtr., p. 18.

C. S. L. Baker
W. S. Fulton
Malaysian Rubber Producers' Research Association

RUBBER, SYNTHETIC.

See ELASTOMERS, SYNTHETIC.

RUBIDIUM AND RUBIDIUM COMPOUNDS

Rubidium [7440-17-7], Rb, is an alkali metal, ie, in Group 1 (IA) of the Periodic Table. Its chemical and physical properties generally lie between those of potassium (qv) and cesium (see CESIUM AND CESIUM COMPOUNDS; POTASSIUM COMPOUNDS). Rubidium is the sixteenth most prevalent element in the earth's crust (1). Despite its abundance, it is usually widely dispersed and not found as a principal constituent in any mineral. Rather it is usually associated with cesium. Most rubidium is obtained from lepidolite [1317-64-2], an ore containing 2–4% rubidium oxide [18088-11-4]. Lepidolite is found in Zimbabwe and at Bernic Lake, Canada.

Rubidium was discovered in 1861 by Bunsen and Kirchhoff by means of an optical spectroscope. It was named for the prominent red lines in its spectrum, from the Latin word *rubidus*, meaning darkest red. Bunsen prepared free rubidium during the same year by an electrolytic method. After cesium, rubidium is the second most electropositive and alkaline element. The two isotopes of natural rubidium are ⁸⁵Rb [13982-12-1] (72.15% o) and ⁸⁷Rb [13982-13-3] (27.85% o). The latter is a beta-emitter having a half-life of 4.9 × 10¹⁰ yr. Twenty-four isotopes of rubidium are known.

A considerable body of literature has been published on the distribution and detection methods of rubidium in geological formations, the oceans, soils, industrial particulate emissions, and stellar interstellar formations (2).

Properties

Physical Properties. Rubidium, a soft, ductile, silvery-white metal, is the fourth lightest metallic element. Having a melting point of 39°C, it can be a liquid at ambient temperatures. Table 1 lists certain physical properties.

Table 1. Properties of Rubidium

Property	Value	References
atomic weight	85.47	
melting point, °C	39.0	3,4
boiling point, °C	689	5,6
density, g cm ³		
solid, 18°C	1.522	3
liquid, 39°C	1.472	3
viscosity, at 39°C, mPa·s(cP)	0.6713	3,7
surface tension at 39°C, mN m(dyn cm)	75	3
vapor pressure, <i>T</i> in K, kPa ^a		
427–1093°C	log ₁₀ <i>P</i> = 3891.8 <i>T</i> + 6.0494	5,8
809–1147°C	log ₁₀ <i>P</i> = 3970.2 <i>T</i> + 9.1546	9
heat of fusion, J g ^b	25.69	10
heat of vaporization, J g ^b	887	1
specific heat, J (kg·K) ^b		
solid	331.37	4
liquid	368.19	5
vapor	241.83	11
thermal conductivity, liquid, W (m·K)	29.3	1
electron work function, eV	2.09	1
critical temperature, <i>T</i> _c , K	1935 ± 87	12
electron affinity, eV	0.486	13,14
neutron-absorption cross-section, thermal, m ^{2c}	7.3 × 10 ⁻²⁹	1
ionization potential, V	4.159	15
	4.1792	14
electronegativity	0.82	14
ionic radius, pm	148	16

^a To convert log₁₀ *P*_{kPa} to log₁₀ *P*_{mm Hg}, add 0.8751 to the constant.

^b To convert J to cal, divide by 4.184.

^c To convert m² to barn, multiply by 10²⁸.

Chemical Properties. The reactions of rubidium are very similar to those of other alkali metals, especially potassium and cesium. Rubidium burns with a violet flame in the presence of air and reacts violently with water, liberating hydrogen which spontaneously explodes if oxygen or air are present. Like other alkali metals, rubidium reacts vigorously with lower alcohols, and violently with halogens, oxidizing agents, and chlorinated hydrocarbons. It also reacts with Teflon. Four oxides of rubidium are known: the yellow rubidium monoxide [12509-27-2], Rb₂O; the dark brown rubidium peroxide [23611-30-5], Rb₂O₂; the black rubidium trioxide [12137-26-7], Rb₂O₃; and the dark orange rubidium superoxide [12137-25-6], RbO₂ (3). Commercial rubidium oxide is usually a mixture of these four oxides. Next to cesium, rubidium is the second strongest Lewis base.

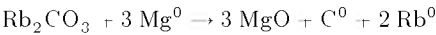
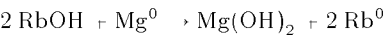
Rubidium metal alloys with the other alkali metals, the alkaline-earth metals, antimony, bismuth, gold, and mercury. Rubidium forms double halide salts with antimony, bismuth, cadmium, cobalt, copper, iron, lead, manganese, mercury, nickel, thorium, and zinc. These complexes are generally water insoluble and not hygroscopic. The soluble rubidium compounds are acetate, bromide, carbonate, chloride, chromate, fluoride, formate, hydroxide, iodide,

nitrate, and sulfate. These compounds are generally hygroscopic. The chemistry of rubidium is essentially restricted to ionic Rb(I) compounds. Although rubidium can form a number of chelates with ether, thiol, and amine ligands, metal–carbon bonds are essentially nonexistent.

Manufacture and Processing

Rubidium is found widely dispersed in potassium minerals and salt brines. Lepidolite [1317-64-2], a lithium mica having the composition $\text{KRbLi}(\text{OH,F})\text{Al}_2\text{Si}_3\text{O}_{10}$, contains up to 3.5% Rb_2O and is the principal source of the element. An ore that is basically pollucite [1308-53-8], $\text{Cs}_2\text{O} \cdot \text{Al}_2\text{O}_3 \cdot 4\text{SiO}_2$, contains up to 1.5% RbO_2 , and some rubidium is produced as a by-product of cesium manufacture from this source. The traditional processes used to recover rubidium from these sources involve extraction of mixed alkali alums from the ore. The ore is subjected to a prolonged sulfuric acid leach to form the alkali alums. The alum solution is filtered from the residue, which is washed with water. Calcination of the ore before leaching increases the yield. The other alkali alums are separated by fractional recrystallizations. The purified rubidium alum is converted to rubidium hydroxide by neutralization to precipitate the aluminum. Subsequent treatment with barium hydroxide precipitates the sulfate. The chlorostannate method requires a partial separation of potassium (16). The remaining dissolved carbonates are converted to chlorides, and the solution is treated with enough stannic chloride to precipitate cesium chlorostannate, which is less soluble than its rubidium counterpart. The cesium-free chloride solution is treated with an excess of stannic chloride to precipitate rubidium chlorostannate [17362-92-4]. The purified rubidium chlorostannate may be decomposed to separate the rubidium and tin chlorides by pyrolytic, electrolytic, or chemical methods. Rubidium compounds can also be separated from other alkali metal compounds by solvent extraction and ion exchange (qv).

Pure rubidium metal is obtained by reducing pollucite or lepidolite ore using an active metal, followed by vacuum distillation (17). Another method is to reduce pure rubidium compounds thermochemically according to the following reactions (18):



Although rubidium is more electropositive than either calcium or magnesium, the equilibrium is driven to the right because the rubidium is continuously distilled away from the reaction mixture. Rubidium metal can be purified by vacuum distillation.

Packaging, Shipping, and Storage. Owing to the reactivity of rubidium metal to oxygen and moisture, the material is considered hazardous and requires special packaging and shipping procedures. The following labeling and shipping restrictions are required in order to be in compliance with the U.S. Department of Transportation (domestic) and the International Air Transport Association (IATA) (international) (19).

Material	Label
rubidium metal	Dangerous When Wet Shipping Class: 4.3 (IATA) Forbidden Passenger Cargo 15-kg Maximum Air Freight
rubidium hydroxide	Corrosive Solid Shipping Class: 8 (IATA) 15-kg Maximum Passenger Cargo 50-kg Maximum Air Freight
rubidium nitrate	Oxidizer
rubidium oxide	Corrosive Solid
rubidium perchlorate	Oxidizer

Special containers are used for storing and shipping rubidium metal (20). For quantities up to 100 g a borosilicate ampul may be used. For larger quantities, stainless steel containers are used to facilitate safe handling. These containers are usually designed to provide a delivery valve, fitted with a dip tube, and an inert gas purge valve. The metal can be easily liquefied and pressurized out of the cylinder using an inert gas such as argon or helium. Both glass and stainless steel containers are hermetically sealed. The metal may also be handled and stored safely in a dry, saturated hydrocarbons within a suitable container.

For shipping purposes, the ampul is placed in aluminum foil or polyethylene bags. The wrapped ampul is packed in a metal can and surrounded by an inert filler material such as vermiculite. These precautions are necessary in order to minimize the chances of ampul breakage during shipment. Most rubidium compounds, however, can be shipped as nonhazardous materials. These compounds are usually stored in glass or polyethylene bottles.

Economic Aspects

The supply and demand for rubidium compounds has grown steadily since the 1970s. In 1979 the U.S. demand was ca 1040 kg of contained rubidium (16), and total world demand was estimated at approximately twice that of the United States. Reserves of rubidium in North America are estimated at 2×10^3 kg; the United States is ~100% import-reliant for rubidium, and Canada is the principal source of the raw material (16). The demand for rubidium metal is small compared to the demand for rubidium compounds. Table 2 lists approximate prices of rubidium metal and rubidium compounds. Primary producers of rubidium are Cabot Performance Chemicals (Boyerstown, Pennsylvania), MSA Research Corporation (Callery, Pennsylvania), and CM Chemical Products, Inc. (Berkeley Heights, New Jersey). Research quantities of rubidium in standard and specialized ampuls are available from Strem Chemicals, Inc. (Newburyport, Massachusetts).

Table 2. Prices of Rubidium Metal and Compounds

Substance	Grade purity ^a , %	
	99 %	99.9 %
metal, \$ / g		
1965–1980	6.50	8.50
1995	10.00	15
compounds, \$ / kg		
1965	64–106	150–172
1980	141–157	227–243
1995	300	350

^a Unless ranges are given, values are approximate.

Specifications and Analytical Methods

Rubidium metal is commercially available in essentially two grades, 99 + % and 99.9 + %. The main impurities are other alkali metals. Rubidium compounds are available in a variety of grades from 99^o to 99.99 + %. Manufacturers and suppliers of rubidium metal and rubidium compounds usually supply a complete certificate of analysis upon request. Analyses of metal impurities in rubidium compounds are determined by atomic absorption or inductive coupled plasma spectroscopy (icp). Other metallic impurities, such as sodium and potassium, are determined by atomic absorption or emission spectrograph. For analysis, rubidium metal is converted to a compound such as rubidium chloride.

Health and Safety Factors

Animal and human toxicity data are available for rubidium compounds (Table 3) (22). Physiological experiments indicate exchangeability of rubidium for potassium in blood, plasma, and tissue. Medical and toxicological literature generally indicate a very low degree of toxicity (23). In many cases, the health risks of rubidium compounds are associated with the anion, eg, hydroxide, or fluoride, rather than the rubidium. Localized ventilation and the use of approved dust respirators are recommended when handling dry rubidium salts, especially hydroxides, oxides, and fluorides. Animal skin tests show that a 5^o RbOH solution is nonirritating to intact skin; however, it is a mild irritant to abraded skin (24). Tests using a 1^o RhOH solution and using a 5^o solution of RbI show no ocular reaction. Concentrated solutions of rubidium hydroxide are severe corrosives to tissues, especially the eyes. A 5^o rubidium iodide solution is not irritating for either intact or abraded skin. As a prudent safety practice, chemical-resistant gloves and approved eye protection are recommended when handling any rubidium salts. Rubidium metal is a severe corrosive to tissue, and should only be handled in an inert atmosphere with complete personal protection.

Table 3. Toxicity Data for Select Rubidium Compounds^a

Compound	CAS Registry Number	LD ₅₀ ^b (oral, rat), mg kg
rubidium carbonate	[584-09-8]	2625
rubidium chloride	[7791-11-9]	4440
rubidium hydroxide	[1310-82-3]	586 ^c
rubidium iodide	[7790-29-6]	4708
rubidium nitrate	[13126-12-0]	4625
rubidium sulfate	[7488-54-2]	4594

^a Ref. 21.
^b LD₅₀ lethal dose, 50^o kill.
^c Threshold limit value (TLV) of RbOH is 0.5 mg m³.

The metal reacts violently with water, ice, steam, lower molecular weight alcohols, and chlorinated hydrocarbons. In the presence of air moisture rubidium can act as an ignition source if a flammable organic liquid or vapor is also present. Rubidium can ignite spontaneously in the presence of oxygen and tarnishes rapidly when exposed to air. Burning rubidium should only be extinguished with dry powders, such as dolomite or sodium carbonate. Some compounds, such as rubidium chromate and rubidium iodide, emit toxic vapors when heated. Other compounds, such as rubidium perchlorate, are irritants.

Rubidium Compounds

Rubidium acid salts are usually prepared from rubidium carbonate or hydroxide and the appropriate acid in aqueous solution, followed by precipitation of the crystals or evaporation to dryness. Rubidium sulfate is also prepared by the addition of a hot solution of barium hydroxide to a boiling solution of rubidium alum until all the aluminum is precipitated. The pH of the solution is 7.6 when the reaction is complete. Aluminum hydroxide and barium sulfate are removed by filtration, and rubidium sulfate is obtained by concentration and crystallization from the filtrate. Rubidium aluminum sulfate dodecahydrate [7488-54-2] (alum), RbAl(SO₄)₂ ·12H₂O, is formed by sulfuric acid leaching of lepidolite ore. Rubidium alum is more soluble than cesium alum and less soluble than the other alkali alums. Fractional crystallization of Rb alum removes K, Na, and Li values, but concentrates the cesium value. Rubidium hydroxide, RbOH, is prepared by the reaction of rubidium sulfate and barium hydroxide in solution. The insoluble barium sulfate is removed by filtration. The solution of rubidium hydroxide can be evaporated partially in pure nickel or silver containers. Rubidium hydroxide is usually supplied as a 50^o aqueous solution. Rubidium carbonate, Rb₂CO₃, is readily formed by bubbling carbon dioxide through a solution of rubidium hydroxide, followed by evaporation to dryness in a fluorocarbon container. Other rubidium compounds can be formed in the laboratory by means of anion-exchange techniques. Table 4 lists some properties of common rubidium compounds.

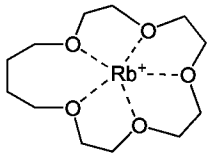
Table 4. Properties of Rubidium Compounds

Compound	CAS Registry Number	Formula	Solubility ^a , g 100 mL		Mp, °C	Bp, °C
			Hot water	Cold water		
rubidium acetate	[563-67-7]	RbC ₂ H ₃ O ₂	86 ^{44 7}		246	
rubidium alum	[7488-54-2]	RbAl(SO ₄) ₂ ·12H ₂ O	43 ⁸⁰	1.3 ⁰	99	
rubidium bromide	[7789-39-1]	RbBr	205 ^{11 3 5}	98 ⁵	682	
rubidium chloride	[7791-11-9]	RbCl	139 ¹⁰⁰	77 ⁰	715	1390
rubidium fluoride	[13446-74-7]	RbF		130.6 ¹⁸	760	1410
rubidium iodide	[7790-29-6]	RbI	163 ²⁵	152 ¹⁷	642	1300
rubidium sulfate	[7488-54-2]	Rb ₂ SO ₄	82 ¹⁰⁰	36 ⁰	1073	
rubidium chromate	[13446-72-5]	RbCrO ₄	95.7 ⁰	62 ⁰		
rubidium nitrate	[13126-12-0]	RbNO ₃	452 ¹⁰⁰	19.5 ⁰		
rubidium carbonate	[584-09-8]	Rb ₂ CO ₃	450 ²⁰		837	740 dec
rubidium hydroxide	[1310-82-3]	RbOH		180 ¹⁵	301	
rubidium perchlorate	[13510-42-4]	RbClO ₄	100 ¹⁸	0.5 ⁰		

^a Superscripted number is temperature in °C.

Owing to the strong electropositive character of rubidium, its chemistry is normally restricted to ionic compounds. The element does not readily

form metal–carbon bonds; therefore, alkyl and aryl compounds of rubidium are generally unstable. A benzylrubidium triammine chelate has, however, been reported (25). The rubidium cation forms a variety of chelated complexes with cyclic ethers, thiols, and amines. The literature is replete with information on alkali metal crown ether complexes (26–32). The stability and formation constant of the crown ether complexes, an example of which is shown, depends on a number of factors, such as the size of the alkali metal, solvent systems, and the size of the crown ether cavity. This chemistry has potential utility in the separation of alkali metals, from each other and from other metals. An area of active research is the organic and organometallic chemistry of fullerenes, particularly buckminster fullerene, C₆₀. C₆₀ is an electron-deficient molecule that can be readily reduced. Rubidium, as do other alkali metals, forms complexes of the type Rb⁺ · xC₆₀[−] where x can be >1. These materials demonstrate, among other properties, superconductivity at low temperatures (33–38).



Uses

The principal use for rubidium metal and its compounds is in research (1,39). Properties are so similar to those of cesium and its compounds that, in many cases, rubidium and cesium may be used interchangeably. The latter, however, is more electropositive and, therefore, usually preferred. In early magnetohydrodynamics (qv) (MHD) and thermionic experiments, rubidium oxide was used (see THERMOELECTRIC ENERGY CONVERSION). When it was used as the seeding material, the conductivity of the plasma improved and the ionization temperature was lowered (see PLASMA TECHNOLOGY). In biological research, rubidium salts, such as RbCl, are used in conjunction with or instead of cesium compounds as a density-gradient medium for ultracentrifugal separation of deoxyribonucleic acid (DNA), viruses, and other large particles (16) (see SEPARATION, CENTRIFUGAL). Rubidium salts are used in pharmaceuticals (qv) as soporifics, sedatives, and in the treatment of epilepsy (see HYPNOTICS, SEDATIVES, ANTICONSULSANTS, AND ANXIOLYTICS). Rubidium iodide has been used as a substitute for potassium iodide in the treatment of goiters. Radioactive rubidium is used as a radioactive tracer for blood flow in the body (40). Additionally, to identify their products in liability and counterfeit cases, manufacturers often add small amounts of rubidium as a chemical tag (16). Cathodes made of rubidium telluride [12210-70-7] are used as radiation detectors for wavelengths in the 200–5000-nm range. Small additions of Rb₂CO₃ reduce the conductivity of glass (qv) and improve its stability and durability. Therefore, Rb₂CO₃ is used in the production of special glasses and in fiber optics (qv). The carbonate is also used in night vision goggles and for synthetic fiber production.

Rubidium-87 emits beta-particles and decomposes to strontium. The age of some rocks and minerals can be measured by the determination of the ratio of the rubidium isotope to the strontium isotope (see RADIOISOTOPES). The technique has also been studied in dating human artifacts. Rubidium has also been used in photoelectric cells. Rubidium compounds act as catalysts in some organic reactions, although the use is mainly restricted to that of a cocatalyst.

BIBLIOGRAPHY

"Rubidium" under "Alkali Metals" in *ECT* 1st ed., Vol. 1, pp. 451–453, by E. H. Burkey, J. A. Morrow, and M. S. Andrew, E. I. du Pont de Nemours & Co., Inc.; "Rubidium Compounds" in *ECT* 1st ed., Vol. 2, pp. 946–949, by J. J. Kennedy, Maywood Chemical Works; "Rubidium and Rubidium Compounds" in *ECT* 1st ed., Suppl. Vol. 2, pp. 734–736, by J. N. Hinyard, American Potash & Chemical Corp.; in *ECT* 2nd ed., Vol. 17., pp. 684–693, by R. E. Davis, American Potash & Chemical Corp.; in *ECT* 3rd ed., Vol. 20, pp. 492–499, by F. B. White, Jr., and W. G. Lidman, Kawecki Berylo Industries, Inc.

- C. A. Hampel, *Rare Metals Handbook*, 2nd ed., Reinhold Publishing Corp., New York, 1961, pp. 434–440.
- "Rubidium: Analysis, Biological Studies and Occurrence," *Chemical Abstracts—Chemical Substances Indexes*, American Chemical Society, Washington, D.C.
- W. Mellor, *Comprehensive Treatise on Inorganic and Theoretical Chemistry*, Vol. 2, Suppl. 3, John Wiley & Sons, Inc., New York, 1963, pp. 2136–2293, 2488–2505.
- "Rubidium," in *Gmelins Handbuch der Anorganischen Chemie*, 8th ed., Vol. 24, Verlag Chemie, Berlin, 1955.
- F. Tepper, A. Murchison, J. Zelenak, and F. Rochlich, *U.S. At. Energy Comm.* ORNL-3605, 26–65 (1963).
- F. M. Perelman, *Rubidium and Cesium*, 1st English ed., The MacMillan Co., New York, 1965.
- R. P. Chhabra and B. K. Srinivas, *Z. Metallkd.* **81**(11), 850 (1990).
- Y. Waseda, S. Ueno, and K. T. Jacob, *J. Mater. Sci. Lett.* **8**(7), 857 (1989).
- A. G. Kalandarishvili, U. K. Mikheev, and D. D. Chilingarishvili, *Teplofiz. Vys. Temp.* **24**(5), 1019 (1986).
- K. K. Kelley, *Bulletin No. 601*, U.S. Bureau of Mines, Washington, D.C., 1962.
- W. D. Weatherford, Jr., J. C. Tyler, and P. M. Ku, *Properties of Inorganic Energy-Conversion and Heat-Transfer Fluids for Space Applications*, WADD Technical Report No. 61-96, Southwest Research Institute, Nov. 1961.
- A. M. Azad and O. M. Sreedharan, *Physica B (Amsterdam)* **153**(1–3), 220 (1988).
- T. A. Patterson, H. Hotop, A. Kasdan, D. W. Norcross, and W. G. Lineberger, *Phys. Rev. Lett.* **39**, 189 (1974).
- S. G. Bratsch, *J. Chem. Educ.* **65**(1), 34 (1988).
- E. N. Simons, *Guide to Uncommon Metals*, Hart Publishing Co., New York, 1967, pp. 161–163.
- R. J. Basile, *Bulletin No. 671*, U.S. Bureau of Mines, Washington, D.C., 1980; *Mineral Commodities Summaries (Rubidium)*, U.S. Geological Survey, Washington, D.C., Jan. 1996.
- U.S. Pat. 3,207,598** (Sept. 21, 1965), C. E. Berthold (to San Antonio Chemicals, Inc.).
- U.S. Pat. 2,668,778** (Feb. 9, 1954), E. A. Taft (to General Electric Co.).
- N. I. Sax, *Dangerous Properties of Industrial Materials*, 9th ed., Van Nostrand Reinhold Publishing Corp., New York, 1996.
- Metals and Alloys—Rubidium*, File No. 320-PD-1, KBI Division of Cabot Corp., Reading, Pa.
- The Registry of the Effects of Toxic Substances*, Department of Health, Education and Welfare, Public Health Services, Center for Disease Control, and the National Institute for Occupational Safety and Health, Washington, D.C., 1977, updated quarterly.
- Kh. Kh. Khamidulina, *Gig. Tr. Prof. Zabol.* (9), 55 (1987); G. Rutigliano, M. Ruggiero, F. Vadrucchio, and S. Lepore, *G. Med. Mil.* **143**(5), 540 (1993).
- U.S. Dept. of Health, Education and Welfare, *Health Hazard Evaluation Report 71-72*, NIOSH, Washington, D.C., July 1973.
- G. T. Johnson, T. R. Lewis, and W. D. Wagner, *Toxicol. Appl. Pharmacol.* **32**, 239 (1975).
- D. Hoffmann and co-workers, *J. Am. Chem. Soc.* **116**(2), 528 (1994).
- Y. Inoue, K. Wada, M. Ouchi, A. Tai, and T. Hakushi, *Chem. Lett.* (6), 1005 (1988).
- A. W. Langer, ed., "Polyamine Chelated Alkali Metal Complexes," *Adv. Chem. Ser.* **130** (1974).
- R. T. Streeter and S. Khazaeli, *Polyhedron*, **10**(2), 221 (1991).
- Y. Inoue, Y. Liu, L. Hui Tong, A. Tai, and T. Hakushi, *J. Chem. Soc. Chem. Commun.* (20), 1556 (1989).
- D. P. Zollinger, E. Bulton, A. Christenhusz, M. Bos, and W. E. Van der Linden, *Anal. Chem. Acta* **198**, 207 (1987).

31.

J. B. Lucas and S. F. Lincoln, *J. Chem. Soc., Dalton Trans.* (4), 423 (1994).

32.

K. Ozutsumi, K. Ohtsu, and T. Kawashima, *J. Chem. Soc., Faraday Trans.* **90**(1), 127 (1994).

33.

W. K. Fullager, I. R. Gentle, G. A. Heath, and J. W. White, *J. Chem. Soc., Chem. Commun.* (6), 525 (1993).

34.

P. Byszewski, R. Jablonski, and S. Kolesnik, *Solid State Commun.* **84**(12), 1111 (1992).

35.

M. Kato and co-workers, *Adv. Supercond. IV, Proc. Int. Symp. Supercond.* **4**, 207–210 (1991).

36.

M. Kraus and co-workers, *Mol. Cryst. Liq. Cryst. Sci. Technol., Sect. A* **245**, 339 (1994).

37.

D. R. Buffinger, S.-M. Lee, R. P. Ziebarth, V. A. Stenger, and C. H. Pennington, *Proc. Electrochem. Soc.* **94-24**, 1044 (1994).

38.

D. W. Murphy and co-workers, *J. Phys. Chem. Solids*, **53**(11), 1321 (1992).

39.

The Economics of Cesium and Rubidium, SWIP #R2, Roskill Information Services, Ltd., London, 1975.

40.

G. L. Brownell, "New Imaging Systems," in *Nuclear Medicine Technical Progress Report*, Massachusetts General Hospital, Boston, Mass., July 1, 1974–Apr. 1, 1975.

General References

J. W. Maustellar, F. Tepper, and S. J. Rodgers, "Alkali Metal Handling and Systems Operating Techniques," *AEC Monograph*, Gordon and Breach, New York, 1967.

C. C. Addison, *The Chemistry of the Liquid Alkali Metals*, John Wiley and Sons, Inc., New York, 1984.

Roland W. Oshe, ed., *Handbook of Thermodynamic and Transport Properties of Alkali Metals*, IUPAC, Blackwell Scientific Publications, Oxford, U.K., 1985.

R. Snaith, *Annu. Rep. Prog. Chem., Sect. A: Inorg. Chem.* **85**, 3 (1988).

"Physics and Chemistry of Alkali Metal Adsorption," in H. P. Bonzel, A. M. Bradshaw, and G. Ertl, eds., *Material Science Monograph*, Vol. 57, Elsevier, New York, 1989.

Frank S. Wagner
Strem Chemicals, Inc.

RUTHENIUM.

See PLATINUM GROUP METALS.

RUTHERFORDIUM.

See ACTINIDES AND TRANSACTINIDES.

SABADILLA, SABADINE, SABALINE.

See INSECT CONTROL TECHNOLOGY.

SACCHARIN.

See SWEETENERS.

SAFETY.

See INDUSTRIAL HYGIENE; MATERIALS RELIABILITY; PLANT SAFETY.

SALICYLIC ACID AND RELATED COMPOUNDS

o-Hydroxybenzoic acid [69-72-7] (salicylic acid) appears very early in the historical recognition and preparation of organic compounds. The extract of roots, bark, leaves, and fruit of many common, widely distributed plants and trees contains the glucosides of methyl salicylate [119-38-6] and salicyl alcohol [90-01-7]. Salicylic acid does not appear in nature in large amounts, but is easily derived from the glucosides by extraction and mild oxidation. As an indicator of early knowledge of the analgesic nature of these materials, Hippocrates, about 400 BC, recommended the extract of willow leaves for relief of pain in childbirth.

Related to salicylic acid as isomers are *m*-hydroxybenzoic acid [99-06-9] and *p*-hydroxybenzoic acid [99-96-7]. The three are commonly known as the monohydroxybenzoic acids; because of the shared chemical reactions, the three isomers are discussed together herein. These monohydroxybenzoic acids have a broad range of applications from paper coatings and liquid crystal preparations to drugs. Salicylic acid and its derivatives are interesting because of their demonstrated activity as analgesics, antipyretics, and antiinflammatory agents, whereas the para-substituted acids and esters known as parabens have activity as preservatives for food.

Physical Properties

o-Hydroxybenzoic acid is obtained as white crystals, fine needles, or fluffy white crystalline powder. It is stable in air and may discolor gradually in sunlight. The synthetic form is white and odorless. When prepared from natural methyl salicylate, it may have a light yellow or pink tint and a faint, wintergreen-like odor. *m*-Hydroxybenzoic acid crystallizes from water in the form of white needles and from alcohol as platelets or rhombic prisms. *p*-Hydroxybenzoic acid crystallizes in the form of monoclinic prisms. Various physical properties of hydroxybenzoic acids are listed in Tables 1–4.

Table 1. Physical Properties of Hydroxybenzoic Acids

Property	Isomer value		
	Ortho	Meta	Para
molecular weight	138.12	138.12	138.12
melting point, °C	159	201.5–203	214.5–215.5
boiling point, °C	211 sub		
density	1.443 ₄ ²⁰	1.473 ₂₆ ²⁵	1.497 ₂₀ ²⁰
refractive index	1.565		
flash point (Tag closed-cup), °C	157		
acid dissociation, <i>K</i> _a , at 25°C	1.05 × 10 ^{–3}	8.3 × 10 ^{–5}	2.6 × 10 ^{–5}
heat of combustion, mJ mol ^a	3.026	3.038	3.035
heat of sublimation, kJ mol ^a	95.14		116.1

^a To convert J to cal, divide by 4.184.

Table 2. Solubilities of the Hydroxybenzoic Acids in Water, Wt %^a

Temperature, °C	Isomer		
	Ortho	Meta	Para
0°C	0.12	0.35	0.25
10°C	0.14	0.55	0.50
20°C	0.20	0.85	0.81
30°C	0.30	1.35	0.81
40°C	0.42	2.0	1.23
50°C	0.64	3.0	2.3
60°C	0.90	4.3	4.2
70°C	1.37	7.0	7.0
80°C	2.21	11.0	12.0

^a Ref. 1.

Table 3. Solubilities of the Hydroxybenzoic Acids in Nonaqueous Solvents, Wt %

Solvent	Isomer ^a		
	Ortho	Meta	Para
acetone at 23°C	396	327	285
benzene at 25°C	0.775	0.010	0.0035
1-butanol	28.8 ₃₈	20.7 _{3 5}	19.5 _{32 5}
ethanol (99 wt % o)	40.6 ₄₁	39.6 ₅	38.75 ₇
<i>n</i> -heptane	2.09 _{92 2}	2.0 ₁₉₇	1.5 ₁₉₇
methanol at 15°C	39.87	40.38	36.22
carbon tetrachloride at 25°C	0.262		
chloroform (satd in H ₂ O) at 25°C	1.84		
ethanol (abs) at 21°C	34.87		
1-propanol at 21°C	27.36		

^a Subscripts are temperature in °C.

Table 4. Saturated Vapor Pressure, *p*, of *o*- and *p*-Hydroxybenzoic Acids

Temperature, °C	<i>p</i> , Pa ^a	
	<i>o</i> -Hydroxybenzoic acid	<i>p</i> -Hydroxybenzoic acid
20°C	0.0	0.0
40°C	0.1	0.0
60°C	1.1	0.0
80°C	8.1	0.0
100°C	46.4	0.3
120°C	222.3	1.9
140°C	909.1	10.6

^a To convert Pa to mm Hg, divide by 133.3.

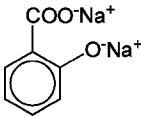
Reactions

The hydroxybenzoic acids have both hydroxyl and the carboxyl groups and, therefore, participate in chemical reactions characteristic of each of these moieties. In addition, these acids can undergo electrophilic ring substitution. The following reactions are discussed in terms of salicylic acid, but are characteristic of all the hydroxybenzoic acids.

Carboxylic Acid Group. Reactions of the carboxyl group include decarboxylation, reduction to alcohols, and the formation of salts, acyl halides, amides, and esters.

Generally, the carboxyl group is not readily reduced. Lithium aluminum hydride is one of the few reagents that can reduce these organic acids to alcohols. The scheme involves the formation of an alkoxide, which is hydrolyzed to the alcohol. Commercially, the alternative to direct reduction involves esterification of the acid followed by the reduction of the ester.

The acid dissolves in aqueous sodium carbonate or sodium bicarbonate to form sodium salicylate [54-21-7]. However, if salicylic acid is dissolved in the presence of alkali metals or excess sodium hydroxide, the disodium compound [13639-21-9] is formed.



Salicylic acid can be converted to salicyloyl chloride [1441-87-8] by reaction with thionyl chloride in boiling benzene. The formation of acyl halide may also extend to reaction with the phenolic hydroxyl. The reaction with phosphorus tri- and pentachlorides is not restricted to the formation of the acid chloride. Further interaction of the phosphorus halide and the phenolic hydroxyl results in the formation of the phosphoric or phosphorous esters.

The formation of amides can be accomplished by dehydration of the ammonium salts of salicylic acid. The more common method for amines is the reaction of the ester, acyl halide, or anhydride with an amine or ammonia. Each step is fast and essentially irreversible.

Esterification is frequently carried out by direct reaction of the carboxylic acid with an alcohol in the presence of a small amount of mineral acid, usually concentrated sulfuric or hydrochloric acid. The esters of commercial importance in both *o*- and *p*-hydroxybenzoic acid are the methyl esters. Direct esterification has the advantage of being a single-step synthesis, but being an equilibrium it is easily reversed. The reaction to the ester is driven by either of

the reactants in large excess or by removal of the reaction product during the course of the reaction. Another esterification method infrequently used is to react the acid chloride with the appropriate alcohol. Still another method is to react the alkali salt with an alkyl or an arylalkyl halide.

Decarboxylation of salicylic acid takes place with slow heating because of the presence of the electronic configuration of the carboxyl group ortho to the hydroxyl group, but does not occur in the other isomers of hydroxybenzoic acid. On rapid heating, salicylic acid sublimes because of its low vapor pressure. This property allows commercial separation from the other isomers as a means of purification analogous to distillation. The differences in the vapor pressures are shown in Table 4.

Hydroxyl Group. Reactions of the phenolic hydroxyl group include the formation of salts, esters, and ethers. The sodium salt of the hydroxyl group is alkylated readily by an alkyl halide (Williamson ether synthesis). Normally, only alkylation of the hydroxyl is observed. However, phenolate ions are ambident nucleophiles and under certain conditions, ring alkylation can also occur. Proper choice of reaction conditions can produce essentially exclusive substitution. Polar solvents favor formation of the ether; nonpolar solvents favor ring substitution.

Esters of the phenolic hydroxyl are obtained easily by the Schotten-Baumann reaction. The reaction in many cases involves an acid chloride as the acylating agent. However, acylation is achieved more commonly by reaction with an acid anhydride. The single most important commercial reaction of this type is the acetylation of salicylic acid with acetic anhydride to produce acetylsalicylic acid [50-78-2] (aspirin).

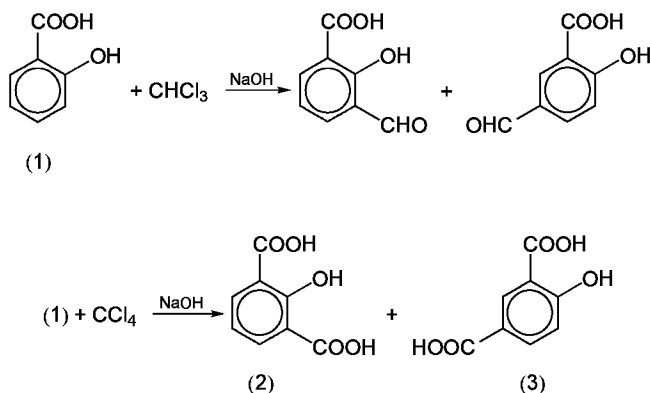
Ring-Substitution Reactions. In the introduction of a third group into a disubstituted benzene, the position the group takes depends on the other groups present. In the case of salicylic acid, the hydroxyl directs ortho- and para-, and the carboxyl directs meta-substitution. The ortho-para director prevails because unlike the meta director it activates the ring. Specifically, the electron-donating hydroxyl group increases the electron density in the 3- and 5-positions. The electron withdrawal nature of the carboxyl group decreases the electron density around the 4- and 6-positions, which further enhances the electron density of the 3- and 5-positions. As a rule, direct substitution occurs more easily in the less sterically hindered 5-position with formation of only small amounts of the 3-substituted and 3,5-disubstituted product. High yields of the 3-substituted salicylic acid usually can only be prepared indirectly.

Direct halogenation of salicylic acid is generally carried out in glacial acetic acid. As expected, the main product is the 5-halo-salicylic acid with small quantities of the 3-halo and 3,5-dihalosalicylic acids.

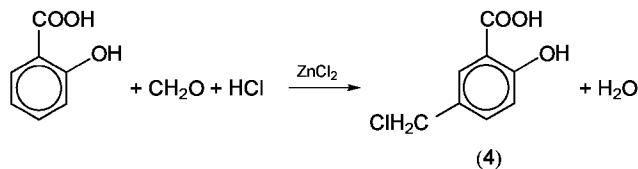
Reaction with cold nitric acid results primarily in the formation of 5-nitrosalicylic acid [96-97-9]. However, reaction with fuming nitric acid results in decarboxylation as well as the formation of 2,4,6-trinitrophenol [88-89-1] (picric acid). Sulfonation with chlorosulfonic acid at 160°C yields 5-sulfosalicylic acid [56507-30-3]. At higher temperatures (180°C) and with an excess of chlorosulfonic acid, 3,5-disulfosalicylic acid forms. Sulfonation with liquid sulfur trioxide in tetrachloroethylene leads to a nearly quantitative yield of 5-sulfosalicylic acid (1).

Because salicylic acid contains the deactivating meta-directing carboxyl group, Friedel-Crafts reactions are generally inhibited. This effect is somewhat offset by the presence of the activating hydroxyl group. Salicylic acid reacts with isobutyl or *t*-butyl alcohol in 80 wt % sulfuric acid at 75°C to yield 5-*t*-butylsalicylic acid [16094-31-8]. In the case of isobutyl alcohol, the intermediate carbonium ion rearranges to (CH₃)₃C⁺.

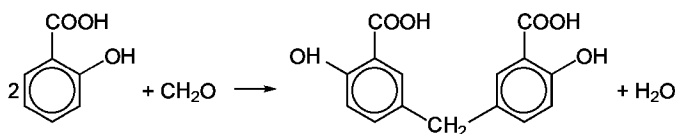
Miscellaneous Reactions. The Reimer-Tiemann reaction of salicylic acid (1) with chloroform and alkali (eq. 1) results in the 3- and 5-formyl derivatives. If the reaction is carried out with carbon tetrachloride, the corresponding dicarboxylic acids form (eq. 2). The products (2) and (3) are 2-hydroxy-1,3-benzenedicarboxylic acid [606-19-2] and 4-hydroxy-1,3-benzenedicarboxylic acid [636-46-4], respectively.



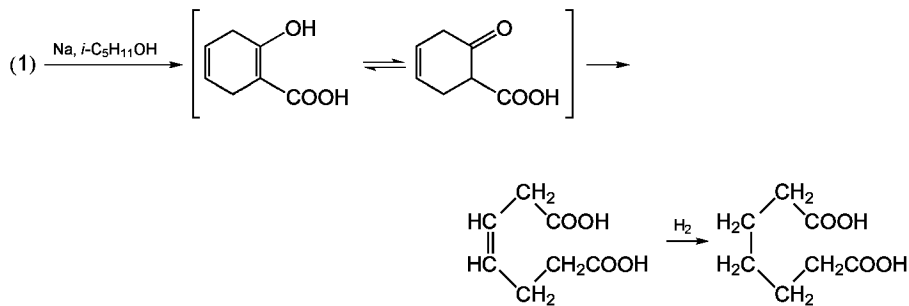
Alkylation involving formaldehyde in the presence of hydrogen chloride is known as chloromethylation (eq. 3). The reagent may be a mixture of formalin and hydrochloric acid, paraformaldehyde and hydrochloric acid, a chloromethyl ether, or a formal. Zinc chloride is commonly employed as a catalyst, although many other Lewis acids can be used. Chloromethylation of salicylic acids yields primarily the 5-substituted product 5-chloromethylsalicylic acid [10192-87-7] (4).



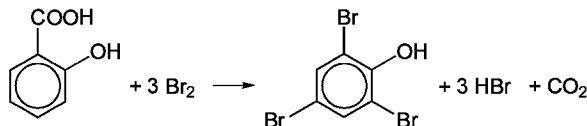
The reaction of salicylic acid with formaldehyde in the presence of catalytic amounts of strong mineral acid results in the condensation product methylene-5,5'-disalicylic acid [122-25-8] (eq. 4).



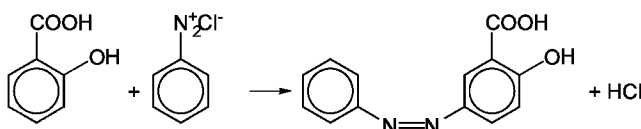
Salicylic acid, upon reaction with amyl alcohol and sodium, reduces to a ring-opened aliphatic dicarboxylic acid, ie, pimelic acid (eq. 5). The reaction proceeds through the intermediate cyclohexanone-2-carboxylic acid. This novel reaction involves the fission of the aromatic ring to *cis*-hexahydrosalicylic acid when salicylic acid is heated to 310°C in an autoclave with strong alkali. Pimelic acid is formed in 35–38% yield and is isolated as the diethyl ester.



During certain substitution reactions, the carboxyl group is often replaced by the entering group. An example is fuming nitric acid, which results in the formation of trinitrophenol. Another is the bromination of salicylic acid in aqueous solution to yield 2,4,6-tribromophenol [25376-38-9] (eq. 6).



Salicylic acid couples with diazonium salts in the expected manner. With diazotized aniline, ie, benzenediazonium chloride, the primary product is 5-phenylazosalicylic acid [3147-53-3] (eq. 7).

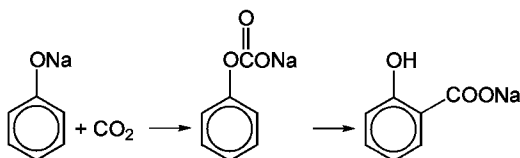


The close proximity of the carboxyl and the hydroxyl groups can be used for heterocyclic synthesis, as in the preparation of hydroxyxanthenes (2).

Manufacture of Salicylic Acid

Natural Occurrence. Salicyl alcohol glucosides [138-52-3] (salicin) occur in *Populus balsamifera* (poplar) and *Salix helix* (willow) trees. Methyl salicylate glucosides occur in *Betula* (birch) and *Fagus* (beech) trees. A more familiar source of methyl salicylate is the leaves of *Gaultheria procumbens* (wintergreen). The color reaction exhibited by iron salts with the salicylates was reported in 1798 (3), which made identification and a crude estimation of the concentration or strength of salicylates in botanical extracts routine for the apothecary. The original name, *Acidium spiricum*, was derived from its identification with one source, the roots, blossoms, and fruit of *Spiraea ulmaria*. During the early history of laboratory chemical synthesis, salicylic acid was first prepared in 1837 through the action of potassium hydroxide on the naturally occurring salicin, a glucoside of salicyl alcohol (4,5).

Early Synthesis. Reported by Kolbe in 1859, the synthetic route for preparing the acid was by treating phenol with carbon dioxide in the presence of metallic sodium (6). During this early period, the only practical route for large quantities of salicylic acid was the saponification of methyl salicylate obtained from the leaves of wintergreen or the bark of sweet birch. The first suitable commercial synthetic process was introduced by Kolbe 15 years later in 1874 and is the route most commonly used in the 1990s. In this process, dry sodium phenate reacts with carbon dioxide under pressure at elevated (180–200°C) temperature (7). There were limitations, however; not only was the reaction reversible, but the best possible yield of salicylic acid was 50%. An improvement by Schmitt was the control of temperature, and the separation of the reaction into two parts. At lower (120–140°C) temperatures and under pressures of 500–700 kPa (5–7 atm), the absorption of carbon dioxide forms the intermediate phenyl carbonate almost quantitatively (8,9). The sodium phenyl carbonate rearranges predominately to the *ortho*-isomer, sodium salicylate (eq. 8).



Current Methods. The general outline of the Kolbe-Schmitt reaction, as it is employed in the 1990s, is as follows. In the first step, phenol and hot aqueous caustic are mixed to produce the sodium phenate which is taken to dryness. Next, the phenate and dry carbon dioxide are introduced into the carbonator. Air is excluded to minimize oxidation and the formation of colored compounds. The gas–solid mixture is agitated and heated, first at low temperature, followed by several hours at higher temperatures, to complete the formation of sodium salicylate. Variations of this reaction have been noted in the literature and are still being investigated (10,11). One reported scheme produces salicylic acid or substituted salicylic acids by reaction of a granulated alkali metal salt of the respective phenolic compound with CO_2 in a fluidized bed at 20–130°C until at least 50–80% of the metal salt has been converted to the corresponding carbonate. In a second step, the temperature of the fluidized bed is increased to 140–210°C to complete the reaction to sodium salicylate or substituted sodium salicylate. These intermediate products are converted to the corresponding acid by acidification with mineral acid. High yields with less interference by tackiness is the claimed advantage (12,13). In most commercial processes, the crude sodium salicylate is cooled and then dissolved in water. If desired, dissolved color bodies can be eliminated via reduction by treating the solution with zinc dust on activated carbon. Finally, the water solution of sodium salicylate is filtered, and then acidified, to precipitate technical-grade salicylic acid. The technical-grade salicylic acid may be further purified by sublimation or recrystallization. During the sublimation and recovery, the risk of dust explosions is minimized by circulating an inert gas through the sublimation chambers.

The carbonation of dried sodium phenate to sodium salicylate, followed by reaction with gaseous hydrogen chloride in a simultaneous neutralization–sublimation has been patented (14). The identified advantages are the minimization of the amount of heat needed for the sublimation process and the elimination of water during acidification, thus saving wastewater cleanup in this process. The heat liberated during the acidification provides the energy needed to sublime the salicylic acid because the heat of acidification and the heat of sublimation are opposite and nearly equal. The subliming chamber in this patented process for salicylic acid is much smaller in volume than the traditional sublimation chambers.

A manufacturing process has been described in which the *o*- and *p*-isomer ratio of the aromatic hydroxycarboxylic acid products resulting from the reaction of an alkali metal phenolate with CO_2 in an organic phosphine oxide solvent is controlled. The mole ratio of the organic phosphine oxide to the alkali metal phenolate is varied from 1 to 4, with the higher ratio favoring para-substitution. The monovalent phenols may be ring-substituted with a 2–12 carbon alkyl or 6–12 carbon aryl group (15). Other syntheses have been reported. The fusion of 2-sulfo- and 2-halobenzoic acid with alkali, and the oxidation of the cresols, *o*-hydroxybenzyl alcohol, and *o*-hydroxybenzaldehyde result in salicylic acid.

An interesting biochemical method of manufacture is the utilization of bioengineered *Pseudomonad* plasmid (16) or *Pseudomonas stutzeri* (17) in a culture medium to oxidize naphthalene or alkyl-substituted naphthalene. The metabolic oxidation products, unsubstituted or substituted salicylic acid,

respectively, are recovered from the medium. DNA coding sequences in a strain of *Pseudomonad* plasmid are engineered to optimize the microbiological oxidation product yield. *P. stutzeri* 5A is cultured at 10–41°C at pH 6–9, and 25–35°C for 7–10 days. Surfactant is added to disperse the sparingly water-soluble naphthalene compound.

Economic Aspects. Technical-grade and USP-grade salicylic acid are produced in the United States by The Dow Chemical Company and in France, Germany (Leuna), and Brazil by Rh  ne-Poulenc. Smaller plants around the world primarily supply regional demands. The world capacity of salicylic acid is approximately 60,000 metric tons per year. In May 1996, the following United States prices per kilogram of product were reported (18): salicylic acid technical-grade, \$2.05; salicylic acid sublimed technical-grade, \$2.30; and salicylic acid USP-grade, \$2.93.

Specifications and Analysis. Specifications for salicylic acid are given in Table 5. The comparison of USP 23 (20), EP 84 (21), and BP 93 (22) shows the somewhat different requirements of these monographs. In monographed ingredient manufacturing, Current Good Manufacturing Practice (CGMP) should be followed. Typical salicylic acid specifications for technical and technical sublimed are given in Table 6.

Table 5. Salicylic Acid Comparison^a

Property	USP 23	EP 84, BP 93, and PF X
identification		
A	Fe test for salicylates	melting point
B	Fe test for salicylates	ir spectrum
C	Fe test for salicylates	Fe test for salicylates
melting range	158–161°C	no test
melting point	no test	158–161°C
loss on drying	0.5° � max	0.5° � max
assay	99.5–101.0° �	99.0–100.5° �
heavy metals	0.002° � max	20 ppm max
chlorides	0.014° � max	100 ppm max
sulfates	0.02° � max	200 ppm max
residue on ignition ^b	0.05° � max	0.1° � max
clarity and color	no test	clean and colorless (1 g in 10-mL alcohol R)
related compounds		
4-hydroxybenzoic acid	0.1° � max	no test
4-hydroxyisophthalic acid	0.05° � max	no test
phenol	0.02° � max	no test
other	0.05° � max	no test
sum	0.2° � max	no test

^a From *United States Pharmacopeia* (USP) and *European* (EP), *British* (BP), and *French* (PF) *Pharmacopeias*.

^b Sulfated ash tests found in BP and EP are considered equivalent to the USP residue on ignition test, except where noted (19).

Table 6. Typical Salicylic Acid Specifications

Type	Melting range, °C	Assay, ° �	Ash, wt � �	Water, wt � �
technical	157–161	98.5	1.0	0.4
technical sublimed	157–161	99.5	0.1	0.1

Health and Safety Factors. Because salicylic acid is a moderate respiratory irritant, skin and eye contact should be avoided and dust mask protection should be used. General good hygiene and good housekeeping should be incorporated into procedures. If large quantities are to be handled, protective clothing with long sleeves and gauntlets as well as approved dust respirators should be used. A further consideration is that dust concentrations as low as 9 g m^{–2} can ignite. For safe practice, ignition and arcing sources should be recognized and eliminated.

The single-dose oral toxicity of salicylic acid is moderate. The LD₅₀ in rats is 400–800 mg kg. Chemical goggles should be worn when handling salicylic acid because eye contact with the chemical can produce irritation and marked pain. If the eyes should come into contact with salicylic acid or its dusts, they should be irrigated promptly and continuously with water for at least 15 minutes before medical attention is sought. Single short exposures are not expected to cause significant irritation, but prolonged or repeated exposures may cause peeling of the skin, rash, itching, or blisters. If the acid contacts the skin, it should be washed off the skin promptly with soap and water. Contaminated clothing should be washed before reuse. If a rash appears, the patient should seek medical attention. Breathing of dusts generated in handling salicylic acid should be avoided. Handlers should wear approved air-purifying respirators. If the product is handled with reasonable care and dusts are controlled by ventilation, there is little likelihood of injury resulting from inhalation.

Uses of Salicylic Acid

During the late 1980s and early 1990s, approximately 60°   of the salicylic acid produced in the United States has been consumed in the manufacture of aspirin [50-78-2]. However, several applications of the acid are growing while the production of aspirin is plateauing, so the proportion is expected to change gradually. Salicylate esters and salts have consumed 20°   of salicylic acid production for a variety of applications, including U.S. FDA-regulated sunscreens and medicaments. Phenolic resins have consumed about 10°   of the production; miscellaneous uses, 10°   (23).

Salicylic acid USP, EP, and other pharmacopeial grades are used medically as antiseptic, disinfectant, antifungal, and keratolytic agents. Salicylic acid is formulated in lotion or ointment formulations for the treatment of dandruff, eczema, psoriasis, and various parasitic skin diseases. Because the keratolytic property of this aromatic acid has use in the safe removal of dead skin cells from the surface of healthy skin, the acid is used in concentrated salicylic acid solutions or suspensions to remove warts and corns. In more dilute form, salicylic acid preparations have found use in dandruff and eczema treatment. Salicylic acid has been considered and found effective by the Advisory Committees to the FDA in various over-the-counter (OTC) drug regulated uses. Among these are acne products, dermatitis, dry skin, dandruff and psoriasis products, and foot care products (24).

Carbonless copypaper using salicylic acid and alkyl salicylic acid derivatives has become an active application area. The salicylic acid is incorporated into a resin or is encapsulated as one of the agents for pressure or thermally activated imaging. The pressure-sensitive dyes used are fluorane and phthalide compounds such as crystal violet lactone, malachite green lactone, benzoyl leuco methylene blue, Japanese pink, dibenzyl green, and One-Dye Black. The developers are mostly metal salts of alkylphenol resins or salicylic acid derivatives. The trend toward high sensitivity thermal color materials is exemplified by products for high speed fax and high brightness labels. A survey of more recent patents shows that improvements in near-infrared-absorbing color former resins are claimed for agents made by the introduction of zinc in the aromatic hydrocarbon resins, including salicylic acid–benzyl chloride

derivative—styrene condensates, and carbonylated *p*-substituted phenol—mesitylene resins. Phenolic resins containing zinc salicylate successfully eliminate the reddening of black images (25,26). Salicylic acid derivatives form surfactants that can influence the rheologic properties of solutions. They impart controllable and useful viscous and elastic properties to aqueous liquids (27).

Several patents of interest include the use of salicylic acid with a quaternary ammonium salt of a long-chain alkyl compound as drag-reducing agents in aqueous media being transported under turbulent flow conditions. One stated application is in hot-water heating systems in large buildings. The large surfactant molecules reduce drag at low (0.2% and lower) concentrations and some of the compounds retain their activity for a year or more even above 90°C (28,29). Salicylic acid has been used as a cross-linking agent in the phenol–formaldehyde resin used in glue for plywood to give faster cures and in the same phenol– formaldehyde resins as a sand core and mold binder imparting higher tensile strength. Salicylic acid is also used as an intermediate in the manufacture of dyes. It is a coupling agent for azo dyes (qv) and a chelating agent in chromium dyes.

Salts of Salicylic Acid

A large number of salts of salicylic acid have been prepared and evaluated for therapeutic or other commercial use. Table 7 lists those most frequently referenced. Sodium salicylate has analgesic, antiinflammatory, and antipyretic activities and was used extensively in the sixteenth and seventeenth centuries as a remedy, prepared from natural sources, for arthritis and rheumatism. In the 1990s the salt can be obtained directly from Kolbe-Schmitt carboxylation or by the reaction of salicylic acid with either aqueous sodium bicarbonate or sodium carbonate. The resulting mixture is heated until effervescence stops; the salt is then isolated by filtration and evaporation to dryness at low temperatures. Generally, the solution must be kept slightly acidic so that a white product is obtained; if the mixture is basic, a colored product results. The USP product contains 99.5–100.5% NaC₇H₅O₃ (anhydrous). The May 1996 price was \$8.15 /kg (18).

Table 7. Physical Properties of Salicylic Acid Salts

Salt	CAS Registry Number	Formula	Mol wt	Crystalline form
bismuth salicylate, basic	[5798-98-1]	Bi(C ₇ H ₅ O ₃) ₃ ·Bi ₂ O ₃	877.3	white microcrystal
calcium salicylate dihydrate	[824-35-1]	Ca(C ₇ H ₅ O ₃) ₂ ·2H ₂ O	350.34	white octahedral
lithium salicylate	[552-38-5]	LiC ₇ H ₅ O ₃	144.05	white powder
magnesium salicylate tetrahydrate	[18917-89-0]	Mg(C ₇ H ₅ O ₃) ₂ ·4H ₂ O	370.60	slightly red crystalline powder
potassium salicylate	[578-36-9]	KC ₇ H ₅ O ₃	176.22	white powder
sodium salicylate	[54-21-7]	NaC ₇ H ₅ O ₃	160.11	white crystalline powder

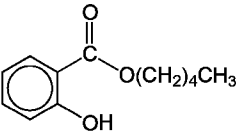
Magnesium salicylate, an analgesic and antiinflammatory agent, appears to have exceptional ability to relieve backaches. It is also used for the symptomatic relief of arthritis. Magnesium salicylate is prepared in a similar manner as sodium salicylate, using the appropriate magnesium carbonate salt. Bismuth subsalicylate [5798-98-1] is employed as an antidiarrheal agent. It is taken orally in combination with other ingredients for protective, antacid action as well as antidiarrheal and antiseptic effects. The price in May 1996 was \$43.30 /kg (18). Other salts of salicylic acid that are of interest are aluminum salicylate [18921-11-4], ammonium salicylate [528-94-9], calcium salicylate, lead salicylate [15748-73-9], lithium salicylate, mercury salicylate [5970-32-1], potassium salicylate, and strontium salicylate [526-26-1]. In the past, many of these salicylates have been used in medicinal applications. These properties as pain-relieving agents have been reviewed by the Internal Analgesic Panel of the FDA using the established advisory panel procedure for the Tentative Final Monograph for Internal Analgesics (30). Only magnesium salts of salicylic acid and sodium are judged safe and effective (Category I). The salts in general also have a few nonmedical applications, eg, uv absorber systems for paints and coatings. However, interest has waned as nonextractable, more efficient uv agents have found their way into commerce. Commercially, the salts of salicylic acid account for less than 5% of the salicylic acid produced.

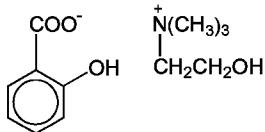
Esters of Salicylic Acid

The esters of salicylic acid account for an increasing fraction of the salicylic acid produced, about 15% in the 1990s. Typically, the esters are commercially produced by esterification of salicylic acid with the appropriate alcohol using a strong mineral acid such as sulfuric as a catalyst. To complete the esterification, the excess alcohol and water are distilled away and recovered. The crude product is further purified, generally by distillation. For the manufacture of higher esters of salicylic acid, transesterification of methyl salicylate with the appropriate alcohol is the usual route of choice. However, another reaction method uses sodium salicylate and the corresponding alkyl halide to form the desired ester.

The main commercial applications for salicylate esters are as uv sunscreen agents and as flavor and fragrance agents. Several have application as topical analgesics. A number of salicylate esters of commercial interest and their physical properties are listed in Table 8.

Table 8. Physical Properties of Salicylic Acid Esters

Ester	CAS Registry Number	R	Mol wt	Mp, °C	Bp, °C ^h Pa	Density g cm ³	Refractive index, <i>n</i> _D	Solubility ^b			
								Water	Methanol	Ethanol	Aqueous
amyl salicylate	[2050-08-0]		208.2		265						
	[]		4								
benzyl	[118-58-1]		228.2	25	320	1.1799 ³⁰ ₄	1.5805 ²⁰	δ	sl	sl	

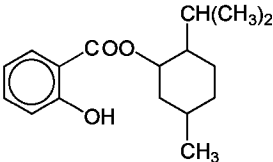


(5)

Phenyl salicylate (salol) is manufactured by heating salicylic acid and phenol in the presence of phosphorus oxychloride for 4–5 hours at 110–115°C. The molten product is separated, mixed with water, dried, and distilled under vacuum. Another process involves the transesterification of a salicylate such as methyl, with phenol in the presence of an alkali or alkaline-earth phenate. Medicinally, phenyl salicylate was formerly used as an intestinal antiseptic. However, the main applications of phenyl salicylate have been related to the ability to absorb uv light over the wavelengths of 290–325 nm. As an effective uv-light absorber, phenyl salicylate was incorporated in alkyd paints, waxes, and polishes, but has been largely replaced in this application by less extractable, more effective compounds. The May 1996 price was \$10.60 kg (18).

Benzyl salicylate can be prepared by the reaction of benzyl chloride with an alkali salt of salicylic acid at 130–140°C or by the transesterification of methyl salicylate with benzyl alcohol. It is used as a fixative and solvent for nitro musks and as a fragrance for detergents. Benzyl salicylate was priced at \$6.75 kg in May 1996 (18).

Homomenthyl salicylate (homosalate), employed as a sunscreen agent, is on a list of 21 compounds for OTC sunscreen products, recommended by the FDA advisory review panel on OTC burn and sunburn prevention as both "safe and effective" (30). Menthyl salicylate (6) is also a sunscreen agent.



(6)

2-Ethylhexyl salicylate [118-60-5] is another compound listed as an accepted sunscreen agent for OTC regulated sunscreen products. The definition of such an agent is "an active ingredient that absorbs 95° or more of the radiation in the uv range at wavelengths from 290 to 320 nm and thereby removes the sunburning rays." By varying the concentration of the agent in the sunscreen vehicle, the sun protection factor (SPF) is developed. It is derived by dividing the exposure response of protected skin by the response of unprotected skin. Standard procedures have been set up to determine the appropriate SPF for these products which are OTC drugs by regulation (31).

Isoamyl salicylate is perhaps the most important ester of salicylic acid for perfumery purposes. Generally, it is manufactured by the transesterification of methyl salicylate. It has a characteristic flowery aroma and is useful in soap fragrances. The May 1996 price was \$5.30 kg (18). Other salicylates of commercial interest as flavor and fragrance agents include isopropyl, isobutyl, phenethyl [87-22-9], and 2-ethylhexyl salicylates.

Salicylsalicylic acid [532-94-3] (salsalate) is prepared by the action of phosphorus trichloride, phosphorus oxychloride, or thionyl chloride on salicylic acid at low temperatures in an appropriate solvent. The crude product is recrystallized rapidly from ethyl alcohol to avoid hydrolysis and esterification. It is used as an analgesic and an antipyretic, as well as in the treatment of acute and chronic rheumatism and arthritis. It does not induce gastric disturbances because it is only slowly hydrolyzed in the intestine. Owing to the slowness of its hydrolysis (two molecules of salicylic acid per molecule of the ester), the action of salicylsalicylic acid is less prompt but more persistent than that of other salicylates. Other salicylates of interest include ethylene glycol monosalicylate [87-28-5], dipropylene glycol monomethylether salicylate, bornyl salicylate [560-88-3], and *p*-acetamidophenyl salicylate [118-57-0].

Other Derivatives of Salicylic Acid

p-Aminosalicylic acid and its salts have been used in the treatment of tuberculosis. *p*-Aminosalicylic acid can be prepared by the carboxylation of *m*-aminophenol (32). Aminosalicylic acid USP assays not less than 98.5° and not more than 100.5°, calculated on the anhydrous basis. The antitubercular agents are likely to be used as the more tolerated salts: calcium [133-15-3], potassium [133-09-5], sodium [133-10-8], and the ethyl [6069-17-2] and phenyl [133-11-9] esters of *p*-aminosalicylic acid.

Methylene-5,5-disalicylic acid is produced by heating two parts salicylic acid with 1–1.5 parts of 30–40 wt ° formaldehyde in the presence of an acid catalyst (33). The resulting product is a mixture of isomers, primarily the 5,5'-isomer and small amounts of low molecular weight polymers. It is used as an intermediate in the production of bacitracin methylenedisalicylate, which is used in a feed supplement to promote growth and as a medicament in swine, feedlot cattle, as well as chickens, turkeys, pheasants, and quail.

Salicylamide [65-45-2] is prepared by the reaction of methyl salicylate with ammonia. Salicylamide has mild analgesic, antiinflammatory, and antipyretic properties. Salicylamide is unlike other salicylates in that it causes sedation and central nervous system depression. Salicylamide is not hydrolyzed to salicylate and its action depends on the entire molecule. Salicylamide has been useful for protection against mildew and fungus in a variety of soaps, salves, lotions, and oils. The May 1996 price was \$8.00 kg (18).

Salicylanilide [87-17-2] is prepared by heating salicylic acid and aniline in the presence of phosphorus trichloride (34). It is used as an intermediate in the production of other chemicals and as a slimicide, fungicide, and medicament. As an active fungicide and antimildew agent, it is used in cotton fabrics, cordage, paints, and lacquers. It is also reported to be a fungistatic agent for plastics. As a medicament, salicylanilide is an active ingredient in creams used to treat fungus infections of the scalp. 3,4,5-Tribromosalicylanilide is an antimicrobial with reported application in soaps, shampoos, textiles, melamine–formaldehyde and polyethylene plastics, synthetic fibers, paints, adhesives, and paper.

5-Sulfosalicylic acid is prepared by heating 10 parts of salicylic acid with 50 parts of concentrated sulfuric acid, by chlorosulfonation of salicylic acid and subsequent hydrolysis of the acid chloride, or by sulfonation with liquid sulfur trioxide in tetrachloroethylene. It is used as an intermediate in the production of dyestuffs, grease additives, catalysts, and surfactants. It is also useful as a colorimetric reagent for ferric iron and as a reagent for albumin. Table 9 shows the physical properties of salicylic acid derivatives.

Table 9. Physical Properties of Salicylic Acid Derivatives

Derivative	CAS Registry Number	Mol wt	Mp, °C	Solubility ^a				
				Water	Methanol	Ethanol	Acetone	Benzene
<i>p</i> -aminosalicylic acid	[65-49-6]	153.14	150–151 dec	sol	sol	sol	sol	insol
methylene-5,5-disalicylic acid	[27496-82-8]	228.26	243–244	δ	sol	sol	sol	δ
salicylamide ^b	[65-45-2]	137.13	142	δ	sol	δ		

salicylanilide ^c	[87-17-2]	213.24	135.8–136.2	sol	δ	δ	δ
5-sulfosalicylic acid	[97-05-2]	218.18	120 (anhydrous)	v	v	v	

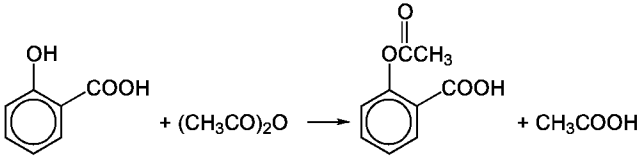
^a δ = partially soluble; v = very soluble.
^b Bp at 14 kPa (105 mm Hg) = 181.5°C; density = 1.175⁴⁰.
^c Also partially soluble in chloroform.

Acetylsalicylic Acid (Aspirin)

Acetylsalicylic acid [50-78-2] (*o*-acetoxybenzoic acid) was first synthesized in 1853 by reaction of acetyl chloride with sodium salicylate. As a drug, acetylsalicylic acid was introduced in Germany in 1899 and into the United States in 1900. The first U.S. patent (35) for the manufacture of acetylsalicylic acid expired in 1917. Aspirin is a registered trademark of Bayer in many nations, but in the United States and the United Kingdom, aspirin is accepted as the generic name for acetylsalicylic acid (36).

Physical Properties. Aspirin normally occurs in the form of white, flat platelets or needle-like crystals, or as a crystalline powder. It melts at 135–137°C and decomposes at 140°C. The solubility of aspirin is about 1 g/300 mL of water at 25°C, about 1 g/5 mL of ethanol at 25°C, 1 g/3.5 mL of acetone at 20°C, and 1 g/17 mL of chloroform at 25°C.

Manufacture. Aspirin [50-78-2] is manufactured by the acetylation of salicylic acid with acetic anhydride (eq. 9) (37,38).



Salicylic acid and acetic anhydride are introduced into a glass-lined or stainless steel reactor. The reactor temperature is kept below 100°C for two to three hours to complete the esterification. The resulting solution is pumped through a filter to remove extraneous solids that interfere with the crystallization, and then into a crystallizer vessel having a reliable temperature control. The temperature is gradually reduced to induce crystallization of the aspirin from the acetic acid/acetic anhydride mother liquor. The resulting suspension is centrifuged. The crystals are washed with water, dried to less than 0.5% by weight moisture, and then separated by particle size. In some processes, the crystals from the centrifuge are recrystallized prior to washing.

Various processes involve acetic acid or hydrocarbons as solvents for either acetylation or washing. Normal operation involves the recovery or recycle of acetic acid, any solvent, and the mother liquor. Other methods of preparing aspirin, which are not of commercial significance, involve acetyl chloride and salicylic acid, salicylic acid and acetic anhydride with sulfuric acid as the catalyst, reaction of salicylic acid and ketene, and the reaction of sodium salicylate with acetyl chloride or acetic anhydride.

Production and Economic Aspects. Aspirin is produced in the United States by The Dow Chemical Company, Rhône-Poulenc, and Norwich (a division of Proctor & Gamble). Globally, Rhône-Poulenc has additional production facilities in France and in Thailand. Bayer is self-supplied from production units in Spain and Turkey; over the years many small plants have been established around the world for regional or country supply. The aspirin market is increasingly globally supplied. Aspirin is generally considered mature, and only population increases and new uses will affect its production and demand, which is thought to be of the order of 30–35,000 t for total world consumption. The May 1995 price was \$8.30/kg (18).

Specifications. Table 10 shows the tests required to satisfy the demands of the *Pharmacopeias* of the United States, Europe, and Japan (20,21,22).

Table 10. Acetylsalicylic Acid Comparison^a

Property	USP 23	EP 84, BP 93, and PF X
assay	99.5–100.5%	99.5–101.0%
identification		
A	color test	ir spectrum
B	ir spectrum	hydrolyzation, mp of salicylic acid
C	no test	W/Ca(OH) ₂ , color test
D	no test	color test
instantaneous melting point ^b	no test	~143°C
loss on drying	0.5% max (silica gel)	0.5% max (vacuo)
readily carbonizable substances	passes Q color	no test
residue on ignition ^c	0.05% max	0.1% max
clarity of solution	clear (500 mg in 10-mL Na ₂ CO ₃ solution)	clear (1 g in 10-mL alcohol R)
color of solution	no test	colorless (1 g in 10-mL alcohol R)
chloride	0.014% max	no test
sulfate	0.04% max	no test
heavy metals	0.001% max	20 ppm max
free salicylic acid	0.1% max	500 ppm max
related substances	no test	0.1% max as acetylsalicylsalicylic acid
organic volatile impurities	meets requirements	no test

^a From *United States Pharmacopeia* (USP) and *European* (EP), *British* (BP), and *French* (PF) *Pharmacopeias*.
^b This is a product character, not a specification.
^c Sulfated ash tests found in BP and EP are considered equivalent to the USP residue on ignition test, except where noted (19).

Uses. Aspirin has analgesic, antiinflammatory, and antipyretic activity. It is used for the relief of less severe types of pain, such as headache, neuritis, acute and chronic rheumatoid arthritis, and toothache. Aspirin can be purchased in a variety of OTC and prescription dosage forms made and formulated by many companies. Tablets, ie, buffered, plain, or enteric-coated, are the most familiar in the United States, but other forms such as powder and effervescent formulations are of considerable importance in other parts of the world.

There has been a rebirth of interest in aspirin between the 1970s and 1990s as evidence accumulated from a number of clinical trials that aspirin ingestion lowers the incidence of myocardial infarction (39,40), unstable angina (41,42), and stroke (43).

These actions of aspirin are thought to result from its ability to reduce the production of prostaglandin formed by platelet cells without appreciably

affecting the other important functions in these blood factors (44).

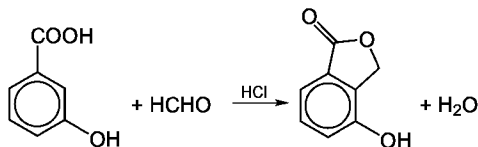
The understanding of these actions of aspirin started in 1971 (45) and resulted in the recommendations of the medical community that small doses of aspirin, used under the care of the doctor, may be a prevention measure for heart attack, and stroke for those considered at risk in the population. Further reported work may expand health benefits to prevention of certain kinds of cancer (46).

m-Hydroxybenzoic Acid

Of the three hydroxybenzoic acids, the metaisomer is of least commercial importance. It offers no outstanding points of chemical interest and is used industrially in small quantities in a limited number of applications.

Reactions. *m*-Hydroxybenzoic acid affords a variety of products, depending on the catalyst and conditions employed. Catalytic reduction over platinum black or platinum oxide in alkaline solution gives 3-hydroxycyclohexanecarboxylic acid [22267-35-2]. Reduction of a warm aqueous solution over platinum oxide or over colloidal platinum yields cyclohexanecarboxylic acid. *m*-Hydroxybenzaldehyde can be prepared by reducing *m*-hydroxybenzoic acid with sodium amalgam. Finally, reduction over Raney nickel gives cyclohexanol.

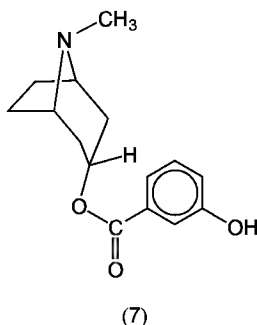
Nitration of *m*-hydroxybenzoic acid with fuming nitric acid in the presence of sulfuric acid and acetic anhydride gives a mixture of the 2-nitro [602-00-6] and 4-nitro [619-14-7] substitution products. Bromination and iodination yield the 4-halogenated derivatives (4-bromo [14348-38-0] and 4-iodo [58123-77-6]). When *m*-hydroxybenzoic acid is treated with formalin in the presence of hydrochloric acid, 4-hydroxyphthalide [13161-32-5] is obtained as shown in equation (10).



Unlike salicylic acid, *m*-hydroxybenzoic acid does not undergo the Friedel-Crafts reaction. It can be converted in 80% yield to *m*-aminophenol by the Schmidt reaction, which involves treating the acid with hydrazoic acid in trichloroethylene in the presence of sulfuric acid at 40°C (47).

Manufacture. *m*-Hydroxybenzoic acid was first obtained by the action of nitrous acid on *m*-aminobenzoic acid (48). It is more conveniently prepared by the sulfonation of benzoic acid with fuming sulfuric acid. The resulting *m*-sulfobenzoic acid is mixed with salt and fused with caustic soda at 210–220°C. The fusion melt is dissolved in water and acidified with hydrochloric acid to precipitate the crude product. Final purification is generally achieved by recrystallization from water.

Uses. *m*-Hydroxybenzoic acid is reported as an intermediate in the manufacture of germicides, preservatives, pharmaceuticals, and plasticizer. In the production of pharmaceuticals, the *m*-hydroxybenzoic acid ester of tropine, ie, (*m*-hydroxybenzoyl)tropeine [52418-07-2] (7), is claimed to cause dilation of the pupils (mydriatic effect), whereas the sodium salt [81256-75-9] is a cholegogic agent promoting the discharge of bile. Esters and metal salts of *m*-hydroxybenzoic acid have been used as germicides and preservatives in foods and meats. Ethers of alkyl esters and carbonates of glycol esters of the acid have been patented as plasticizers for vinyl and cellulosic plastics. It is useful in the manufacture of B-stage epoxy resins having long shelf lives and short curing times (49).



p-Hydroxybenzoic Acid

p-Hydroxybenzoic acid is of significant commercial importance. The most familiar application is the use of several of its esters as preservatives, known as parabens. Also of interest is the use in liquid crystal polymer applications.

Reactions. *p*-Hydroxybenzoic acid undergoes the typical reactions of the carboxyl and hydroxyl moieties. When heated above its melting point, it decomposes almost completely into phenol and carbon dioxide. It reacts with electrophilic reagents in the predicted manner and does not undergo the Friedel-Crafts reaction. Nitration, halogenation, and sulfonation afford the 3-substituted products. Heating *p*-hydroxybenzoic acid with 8 *N*-nitric acid results in a 95% yield of picric acid. In a similar fashion, treatment with chlorine water yields 2,4,6-trichlorophenol (50).

Manufacture. Several methods have been described for the preparation of *p*-hydroxybenzoic acid. The commercial technique is similar to that of salicylic acid, ie, Kolbe-Schmitt carboxylation of phenol. The modification includes the use of potassium hydroxide in place of caustic (51). The dried potassium phenate is heated under pressure, 270 kPa (2.7 atm) or more, with dry carbon dioxide at 180–250°C. The potassium salt [16782-08-4] of *p*-hydroxybenzoic acid forms almost quantitatively and can be converted to free acid by using a mineral acid.

Other reported syntheses include the Reimer-Tiemann reaction, in which carbon tetrachloride is condensed with phenol in the presence of potassium hydroxide. A mixture of the ortho- and para-isomers is obtained; the para-isomer predominates. *p*-Hydroxybenzoic acid can be synthesized from phenol, carbon monoxide, and an alkali carbonate (52). It can also be obtained by heating alkali salts of *p*-cresol at high temperatures (260–270°C) over metallic oxides, eg, lead dioxide, manganese dioxide, iron oxide, or copper oxide, or with mixed alkali and a copper catalyst (53). Heating potassium salicylate at 240°C for 1–1.5 h results in a 70–80% yield of *p*-hydroxybenzoic acid (54). When the dipotassium salt of salicylic acid is heated in an atmosphere of carbon dioxide, an almost complete conversion to *p*-hydroxybenzoic acid results. The *p*-aminobenzoic acid can be converted to the diazo acid with nitrous acid followed by hydrolysis. Finally, the sulfo- and halogenobenzoic acids can be fused with alkali.

Uses. There are many polymer and plastic applications for *p*-hydroxybenzoic acid. A linear polymer of *p*-hydroxybenzoic acid displays long-term stability in air at over 325°C (55). In addition, the polymer has a self-lubricating character combined with a high elastic modulus, thermal conductivity, electrical insulating character, and solvent resistance. These properties make useful metal coatings that are durable. Another application of interest is in liquid crystal preparation. *p*-Hydroxybenzoic acid has been studied in liquid crystal systems either as part of a rod- or disk-shaped surfactant molecule in a solvent, or as a monomer in a polymer, such as a polyester, that exhibits a degree of order lower than the crystalline solid, but higher than the normal isotropic liquid. Such materials are known as liquid crystals because they combine in a single phase the typical flow of a liquid with the anisotropy of properties found in noncubic crystals (56,57,58) (see LIQUID CRYSTALLINE MATERIALS).

p-Hydroxy benzoic acid is used in the manufacture of the methyl, ethyl, *n*-propyl, *n*-butyl, and benzyl esters called parabens. These esters have been used as preservatives for food, pharmaceuticals, and cosmetics for many years. Physical properties are listed in Table 11. These esters are effective bacteriostatic and fungistatic agents against a wide variety of microorganisms. The specifications for the food ingredient chemicals are given in the *Food Chemicals Codex* (59). The compilation of standards for food-grade chemicals was brought together by the Food Protection Committee of the National

Academy of Sciences and the National Research Council effort in the early 1960s to provide standards similar to the USP and NF for drug ingredients. Unlike the USP–NF, the *Food Chemicals Codex* does not have the same legal status as the USP–NF compilation. The exception to this involves definitions and interpretive regulations relating to the eligibility of substances for classification as GRAS (generally recognized as safe) published in 1971.

Table 11. Physical Properties of Alkyl *p*-Hydroxybenzoates (Parabens)

Property	Methyl	Ethyl	<i>n</i> -Propyl	Butyl	Benzyl
CAS Registry Number	[99-76-3]	[120-47-8]	[94-13-3]	[94-26-8]	[94-18-8]
mp, °C	125–128	116–119	95–98	68–72	108–113
assay (min), %	99.0	99.0	99.0	99.0	99.0
solubility, at 25°C, g / 100 g solvent					
water	0.25	0.17	0.05	0.02	0.006
water (at 80°C)	2	0.86	0.30	0.15	0.09
methanol	59	115	124	220	79
ethanol	52	70	95	210	72
propylene glycol	22	25	26	110	13
peanut oil	0.5	1	1.4	5	0.5
acetone	64	84	105	240	102
benzene	0.7	1.65	3	40	2.6
ether	23	43	50	150	52
carbon tetra-chloride	0.1	0.9	0.8	1	0.08

Salicyl Alcohol

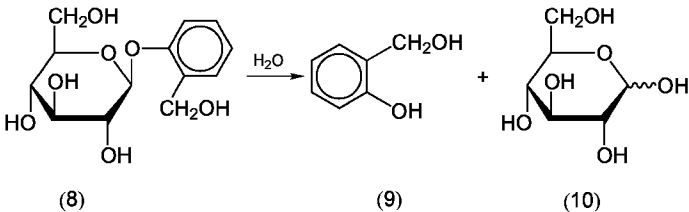
Salicyl alcohol [90-01-7] (saligenin, *o*-hydroxybenzyl alcohol) crystallizes from water in the form of needles or white rhombic crystals. It occurs in nature as the bitter glycoside, salicin [138-52-3], which is isolated from the bark of *Salix helix*, *S. pentandra*, *S. praeox*, some other species of willow trees, and the bark of a number of species of poplar trees such as *Populus balsamifera*, *P. candicans*, and *P. nigra*.

Physical Properties. The alcohol, which sublimes readily and is very soluble in alcohol and ether, has the following properties: melting point, 86°C; density 13° 25°, 1.161 g / cm³; heat of combustion, 3.542 mJ / mol (846.6 kcal / mol); and solubility in 100 mL water at 22°C, 6.7 g.

Reactions. Saligenin [90-01-7] undergoes the typical reactions of phenols and benzyl alcohol. When heated above 100°C, it transforms into a pale yellow resinous material. Amorphous condensation products are obtained when saligenin reacts with acetic anhydride, phosphorus pentachloride, or mineral acids. Upon boiling with dilute acids, saligenin is converted into a resinous body, saliretin, a condensed form of saligenin. Condensation reactions of saligenin with itself in the absence of any catalysts and in the presence of bases have also been studied.

Oxidation of saligenin with chromic acid or silver oxide yields salicylaldehyde as the first product. Further oxidation results in the formation of salicylic acid, which is also obtained when saligenin is heated with sodium hydroxide at 200–240°C. Chlorination of an aqueous solution of the alcohol gives 2,4,6-trichlorophenol, and bromination in an alkaline medium yields 2,4,6-tribromophenol and tribromosaligenin. When saligenin is heated with one mole of resorcinol in the presence of anhydrous zinc chloride, 3-hydroxyxanthene forms.

Manufacture. The hydrolysis of the naturally occurring β-glycoside (salicin) (8) with hydrochloric or sulfuric acid affords saligenin (9) and glucose (10) (eq. 11).



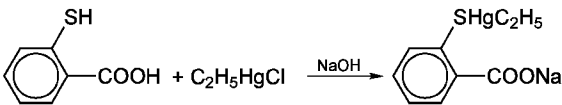
Numerous methods for the synthesis of salicyl alcohol exist. These involve the reduction of salicylaldehyde or of salicylic acid and its derivatives. The alcohol can be prepared in almost theoretical yield by the reduction of salicylaldehyde with sodium amalgam, sodium borohydride, or lithium aluminum hydride; by catalytic hydrogenation over platinum black or Raney nickel; or by hydrogenation over platinum and ferrous chloride in alcohol. The electrolytic reduction of salicylaldehyde in sodium bicarbonate solution at a mercury cathode with carbon dioxide passed into the mixture also yields saligenin. It is formed by the electrolytic reduction at lead electrodes of salicylic acids in aqueous alcoholic solution or sodium salicylate in the presence of boric acid and sodium sulfate. Salicylamide in aqueous alcohol solution acidified with acetic acid is reduced to salicyl alcohol by sodium amalgam in 63% yield. Salicyl alcohol forms along with *p*-hydroxybenzyl alcohol by the action of formaldehyde on phenol in the presence of sodium hydroxide or calcium oxide. High yields of salicyl alcohol from phenol and formaldehyde in the presence of a molar equivalent of ether additives have been reported (60). Phenyl metaborate prepared from phenol and boric acid yields salicyl alcohol after treatment with formaldehyde and hydrolysis (61).

Uses. Saligenin has been used medically as an antipyretic and appears to possess marked topical analgesic powers in concentrations of 4–10%. Saligenin's taste is pungent at first and then numbing. Unsymmetrical diphenylmethanes prepared from salicyl alcohol and substituted phenol at 160–170°C in the presence of alkaline catalysts have been claimed as resin components and resin-hardening agents (62).

Thiosalicylic Acid

Thiosalicylic acid[147-93-3] (*o*-mercaptobenzoic acid), a sulfur-yellow solid that softens at 158°C, has a melting point of 164°C. It sublimes, is slightly soluble in hot water but freely soluble in glacial acetic acid and alcohol, and yields dithiosalicylic acid [527-89-9] upon exposure to air.

Reactions. Thiosalicylic acids reacts with ethylmercuric chloride in alcohol and in the presence of sodium hydroxide to yield sodium ethylmercurithiosalicylate [54-64-8] (thimerosal; Merthiolate, Eli Lilly and Company) (63) (eq. 12).



Uses. Thiosalicylic acid has been used as an anthelmintic, bactericide, and fungicide. It has also been used as a rust remover, a corrosion inhibitor for steel, and a polymerization inhibitor. In photography, it has application in print-out emulsions and as an activator for photographic emulsions.

BIBLIOGRAPHY

"Salicylic Acid, Salicylaldehyde, and Salicyl Alcohol" in *ECT* 1st ed., Vol. 12, pp. 45–66, by R. T. Gottesman, Heyden Chemical Corp.; "Salicylic Acid" in *ECT* 2nd ed., Vol. 17, pp. 720–743, by R. T. Gottesman and D. Chin, Tenneco Chemicals, Inc.; "Salicylic Acid and Related Compounds" in *ECT* 3rd ed., Vol. 20, pp. 500–524, by S. H. Erickson, Dow Chemical Co.

1. E. E. Gilbert and co-workers, *Ind. Eng. Chem.* **45**, 2067 (1953).
2. P. Grover, G. Shah, and R. Shah, *J. Chem. Soc.* 3982 (1955).
3. W. White, *Observations and Experiments on the Broad-Leaved Willow Bark*, Hazard, Bath, U.K., 1798.
4. R. Piria, *Ann. Chim. Phys.* **69**, 281 (1838).
5. R. Piria, *Prakt. Liebigs Ann.* **30** (1839).
6. H. Kolbe and E. Lautemman, *Prakt. Liebigs. Ann.* **115**, 157 (1860).
7. H. Kolbe, *J. Prakt. Chem.* **21**, 443 (1880).
8. **Ger. Pat. 29,939**, R. Schmitt.
9. **Ger. Pat. 38,742**, R. Schmitt; *J. Prakt. Chem.* **23**, 141 (1882).
10. A. S. Lindsey and H. Jeskey, *Chem. Rev.* **57**, 583 (1957).
11. **U.S. Pat. 2,824,892** (Feb. 25, 1958), L. B. Barkley (to Monsanto Chemical Co.).
12. **Eur. Pat. 89565** (Sept. 28, 1983), **Ger. Pat. 3210599** (Oct. 6, 1983), **U.S. Pat. 4529817** (July 16, 1985), H. Karkossa, G. Stopp, and V. Trescher (to Bayer AG).
13. **Eur. Pat. 72095** (Feb. 16, 1983), **U.S. Pat. 4376867** (Mar. 15, 1983), G. Jansen and P. Wolff.
14. **U.S. Pat. 4137258** (Jan. 20, 1979), E. R. Moore and co-workers (to The Dow Chemical Co.); **U.S. Pat. 4,171,453** (Oct. 16, 1979), E. R. Moore and co-workers (to The Dow Chemical Co.).
15. **Eur. Pat. 478197** (Apr. 1, 1992) T. Nakanishi and T. Miura (to Mitsui Toatsu Chem. Inc.).
16. **Jpn. Pat. 59501732** (Oct. 18, 1984), **Eur. Pat. 103887** (Mar. 3, 1984) B.D. Ensesly (to Amgen).
17. **Jpn. Pat. 63207378**, **Jpn. Pat. 63207392** (Aug. 26, 1988), T. Tukasue and O. Foji (to Sumitomo Metal Ind KK).
18. *Chem. Mark. Rep.* (May 31, 1996).
19. *United States Pharmacopeia* 23, **281**, 1731 (1995).
20. *United States Pharmacopeia 23/National Formulary 18*, United States Pharmacopeial Convention, Inc., Rockville, Md., 1995.
21. *European Pharmacopeia*, 2nd ed., Maisonneuve SA, Sainte Ruffine, France, 1984.
22. *British Pharmacopeia 1993*, HMSO, U.K., 1993.
23. "Chemical Profile on Salicylic Acid," *Chem. Mark. Rep.* (Feb. 22, 1993).
24. W. E. Gilbertson, *Handbook of Nonprescription Drugs*, 9th ed., American Pharmaceutical Association, Washington, D.C., 1990, pp. 25–39.
25. M. Matsumoto, *Jpn. J. Pap. Technol.* **4**, 20–24 (Apr. 1990).
26. Y. Tokunaga, *Jpn Tappi J.* **6**, 10–13 (June 1990).
27. G. D. Rose and A. S. Teot, in C. A. Herb and R. K. Prud'homme, eds., *Structure and Flow in Surfactant Solutions*, American Chemical Society, Washington, D.C., 1994, p. 353.
28. **Eur. Pat. 97926** (Jan. 11, 1984), H. Hoffmann and co-workers (to Hoechst AG).
29. **Ger. Pat. 3347378** (July 11, 1985) D. Ohlendorf and co-workers (to Hoechst AG).
30. *Fed. Reg.* **21**, 53, 221 (Nov. 16, 1988).
31. E. M. DeSimone II, *Handbook of Nonprescription Drugs*, American Pharmaceutical Association, Washington, D.C., 1990, p. 909.
32. **U.S. Pat. 2,644,011** (June 30, 1953), R. P. Parker and J. M. Smith, Jr. (to American Cyanamid Co.).
33. **U.S. Pat. 2,774,712** (Oct. 13, 1956) E. L. Baron (to S. B. Penick & Co.).
34. **U.S. Pat. 2,763,683** (Sept. 18, 1956), F. L. Beman and E. C. Britton (to The Dow Chemical Co.).
35. **U.S. Pat. 644,077** (Feb. 27, 1900), F. Hoffman (to Farbenfabriken Bayer and Heyden).
36. C. C. Mann and M. L. Plummer, *The Aspirin Wars*, Harvard Business School Press, Boston, Mass., p. 50.
37. **U.S. Pat. 2,987,539** (June 6, 1961), W. C. Stoesser and W. R. Surine (to The Dow Chemical Co.).
38. **U.S. Pat. 3,235,583** (Feb. 15, 1966), R. T. Edmunds (to Norwich Pharmacal Co.).
39. C. H. Hennekens and co-workers, *Circulation*, **80**, 749–756 (1989).
40. C. H. Hannekens and co-workers, *Lancet*, 349–360 (1988).
41. D. J. Fitzgerald and co-workers, *N. Eng. J. Med.* **315**, 983–389 (1986).
42. H. D. Lewis and co-workers, *N. Eng. J. Med.* **309**, 396–403 (1983).
43. H. J. M. Barnett, *Stroke*, **21**(Suppl. IV), 40–43 (1990).
44. W. L. Smith, *Drug News Perspect.* **4**(6), (July 1991).
45. J. R. Vane, *Postgrad. Med. J.* 743 (Dec. 1983); J. R. Vane, *Nature New Bio.* 232 (June 23, 1971); J. R. Vane, *Hospital Prac.* 61 (Mar. 1972).
46. M. J. Thun, M. M. Namboodiri, and C. W. Heath, *N. Eng. J. Med.* **325**, 1593–1596 (1991).
47. L. H. Briggs and J. W. Lyttleton, *J. Chem. Soc.* 421 (1943).
48. Gerland, *Ann.* **86**, 143 (1853); **91**, 189 (1854).
49. **Ger. Pat. 1,017,699** (Jan. 19, 1966), M. P. Rainton (to CIBA ARI. Ltd.).
50. K. V. Sarkanen and C. W. Dence, *J. Org. Chem.* **25**, 715 (1960).
51. **Can. Pat. 843457** (June 2, 1970), R. Ueno and T. Mayazaki (to Ueno Pharmaceutical Research Laboratory).
52. **Jpn. Pat. 31,333** (Dec. 15, 1969), H. Yasuhara, T. Nogi, and I. Takahashi (to Toyo Rayon Co., Ltd.); H. Yasuhara and T. Nogi, *Chem. Ind.* **77** (1969).
53. **Belg. Pat. 665,631** (Dec. 4, 1965), W. W. Kaeding and E. J. Strojny (to The Dow Chemical Co.).
54. C. A. Buehler and W. E. Cate, in A. H. Blatt, ed., *Organic Syntheses*, Collective Vol., John Wiley & Sons, Inc., New York, 1943, p. 341.
55. J. Economy, B. E. Nowak, and S. G. Cottis, *Polym. Prep.* **2**(1), 332 (1970); *Sample J.* **6**(6), 21 (1970).
56. G. Vertogen and W. H. deJeu, *Thermotropic Liquid Crystals, Fundamentals* Springer-Verlag, Berlin, Germany, 1988.
57. R. A. Weiss and C. K. Ober, ed., *Liquid-Crystalline Polymers*, American Chemical Society, Washington, D.C., 1990.
58. B. Bahadur, ed., *Liquid Crystals, Applications and Uses*, Vol. 2, World Scientific, London, 1990.
59. Committee on Codex Specifications, *Food Chemicals Codex*, 3d ed., National Academy Press, Washington, D.C., 1981.
60. G. Casiraghi and co-workers, *Synthesis*, **2**, 124 (1980).
61. **U.S. Pat. 3,321,526** (May 23, 1967), P. A. R. Marchand and J. B. Grenet (to Rhône Poulenc SA).
62. **U.S. Pat. 2,744,882** (May 8, 1956), H. L. Bender and co-workers (to Union Carbide Corp.).
63. **U.S. Pat. 1,672,615** (1928), M. S. Kharasch; Trikojus, *Nature*, **158**, 472 (1946); Swirski and co-workers, *Przemysl. Chem.* **39**, 371 (1960).

General References

M. Gross and L. A. Greenberg, *The Salicylates: A Critical Bibliographic Review*, Hillhouse Press, New Haven, Conn., 1948.
M. J. H. Smith and P. Smith, *The Salicylates: A Critical Bibliographic Review*, Wiley-Interscience, New York, 1966.
K. D. Rainsford, *Aspirin and the Salicylates*, Butterworths, London, 1984.

Hydroxycarbonsauren, *Beilsteins Handbuch der Organischen Chemie*, Springer-Verlag, Berlin, Germany, 1927.
H. Stephen and T. Stephen, *Solubilities of Inorganic and Organic Compounds*, Vol. 1, Parts 1 and 2, Macmillan Publishing, New York, 1963.
J. Timmermans, *Physico-Chemical Constants of Pure Organic Compounds*, Vol. 2, Elsevier Science Publishing Co., Inc., New York, 1965.

Mary R. Thomas
The Dow Chemical Company



SAMARIUM.

See LANTHANIDES.

SAMPLING

The chemical industry produces material in the gaseous, liquid, and solid states ranging from basic chemicals to functional specialties. In addition, many processes require the use of intermediates which may or not be in the physical form of the products. These various materials are sampled for the purpose of process control (qv), product quality control (see QUALITY ASSURANCE), environmental control, and occupational health control. Samplings of some of these materials can be hazardous, particularly those involving toxic, unstable, or pressurized substances, which require special safety precautions. However, excluding these special circumstances, the main problems encountered in sampling materials in the chemical industry are those of selecting an appropriate sampling procedure and hence sampling device to obtain a representative sample. Experience involving a large variety of materials has produced many methods of sampling, only some of which are discussed herein. Extensive coverage of sampling methods is available from the ASTM (1). A summary of some of the ASTM sampling standards is given in Table 1.

Table 1. ASTM Sampling Standards

Material	ASTM standard	Material	ASTM standard
aerospace components	F304	graphite electrodes	C783
		hydrocarbon fluids	D3700
		ion-exchange materials	D2687
		iron ore	E877
			D2813
fluids	F301, F302, F303, F306, F309, F310, F329	leather and leather products	
aggregates	D75	lime and limestone	C50
sample reduction	F306		D2755
	C602	magnesium oxide, electrical-grade	
agricultural liming materials			B215
alkylbenzenesulfonates	D1568	metal powders, finished lots	
	E716		B602
aluminum and aluminum alloys		metallic coatings, electrodeposited	
	D480	metallographic specimens	E3
aluminum powder and paste			E88, E55
asbestos		metals and alloys, nonferrous	
airborne	D4240		E173
amphibole	D3879	metals, for chemical analysis	
cement flat sheets	C459	mica paper	D1677
chrysotile	D2590	microspheres, hollow	D2841
atmospheric analysis			D3438
gases and vapors	D1605		
organic compounds	D3686	naphthalene, maleic anhydride, phthalic anhydride	
bituminous materials	D140		D1466
bituminous mixtures	D5361		
	D3665	oils, drying; fatty acids and polymerized fatty acids	
bituminous paving mixtures		paints	D3925
	D3394	paper	D202
board, electrical insulating			D585
	D2022		
bleaches containing chlorine		paper, paperboard, fiberboard, and related products	
brick	C67		
calcium chloride	D345		
carbon black	D1900, D1799	particulate matter from stacks and flues, for emission testing	D2928, D3685, D4536
cellulose pulps	D3376	peat materials	D2944
cement, hydraulic	C183	pesticides	F725
	C163		D270, D1265
cements, thermal insulating		petroleum and petroleum products	
chemicals, industrial	E300	petroleum products	D4177, D4057
clay tile, structural	C67	petroleum, liquefied	D1265
		phenol	D3852
		pine oil	D802
		pine tars and pine tar oils	D856
coal	D388, D410, D2013, D2022, D2234, D4596, D4596	pinene	D233
pulverized	D197	pitch	D4296
	D346	plasticizers	D1045

coke, for laboratory analysis		plastics	D1898
concrete	C140	resins, lacquers	D29
	C823	rosin	D509
hardened in construction		rubber	
	C172		D1485, D3896, D3138
concrete test, freshly mixed		rubber, raw	D1485
copper alloys	E255	sand, standard	C778
cotton fibers	D1441	shellac varnish	D1650
	D38	soap and soap products	D460
		soil	
creosote and creosote-coal-tar solution		by auger borings	D1452
creylic acid	D3852	by ring-lined barrel	D3550
cyclic products, liquid	D3437	by split-barrel sampler	D1586
	D501	by thin-walled tube	D1587
detergents, inorganic alkaline			
	D801	soil and rock for engineering purposes	D2113, D420, E311
dipentene and related terpane solvents		solvents, volatile	D268
	D923	steam	D1066
electrical insulating liquids			C390
	E32	thermal insulation, preformed	
ferroalloys for size	A610	turpentine	D233
fiberboard	D585	uranium hexafluoride	C1052
fibers, synthetic staple	D3333	uranium ore	C1075
gas		varnishes, insulating	D4733
from transformer	D3305, D2759		D3370, D1192
manufactured	D1247	water microbiological	F1094
natural	D1145	wool	
	F307		D1060, D2525, D584,
			D1234
pressurized from aerospace systems		yarn	D2258
gases, toxic	D4490		
glass containers	C224		
glycerin	C783, D1258		

Definitions and Problems

Sampling is the operation of removing a portion from a bulk material for analysis in such a way that the portion removed has representative physical and chemical properties of that bulk material. From a statistical point of view, sampling is expected to provide analytical data from which some property of the material may be determined. These data should have known and controlled errors and be produced at low cost.

For pure liquids and gases, sampling is relatively easy. Sampling of these media becomes difficult, however, when particulates are involved. Almost all samples taken in the chemical industry contain solids in some form. Some raw materials, intermediates, or products are themselves particulates; others contain particulates as contaminants. Some materials exist in natural deposits, eg, strata, in a heap on the ground; in storage tanks, bins, pipes, or ducts; in railcars, drums, bottles, and bales; or in other containers that may or may not be subdivided easily into representative units. Furthermore, many systems containing particulates tend to segregate during handling or storage, and this may introduce a sampling error in the form of bias. In a heap of particulate material that exhibits segregation, coarse particles collect at the bottom perimeter of a heap, whereas fine particles concentrate in the center core. For particulate systems that segregate in this manner, the act of pouring the material onto a stationary apex to form a heap, fill a bin, load a conveyor, etc, always results in this type of cross-sectional pattern. Thus, any form of sampling from such a stationary heap is biased with respect to particle size distribution and should be avoided. Because the distribution of chemical components in the material may be size-dependent, even a chemical assay from a sample obtained in this way may be in error.

Generally, little is known in advance concerning the degree of homogeneity of most sampled systems. Uniformity, rarely constant throughout bulk systems, is often nonrandom. During the production of thousands of tons of material, size and shape distribution, surface and bulk composition, density, moisture, etc, can vary. Thus, in any bulk container, the product may be stratified into zones of variable properties. In gas and liquid systems, particulates segregate and concentrate in specific locations in the container as the result of sedimentation (qv) or flotation (qv) processes.

Figure 1 shows the particulate loading of a pipe containing gas and particulates where the nonuniformity induced by a disturbance, ie, a 90° bend, is obvious (2). A profile of concentration gradients in a long, straight, horizontal pipe containing suspended solids is shown in Figure 2. Segregation occurs as a result of particle mass. Certain impurities, eg, metal-rich particulates, however, occur near the bottom of the pipe; others, eg, oily flocculates, occur near the top (3). Moreover, the distribution may be affected by liquid-velocity disturbances and pipe roughness.

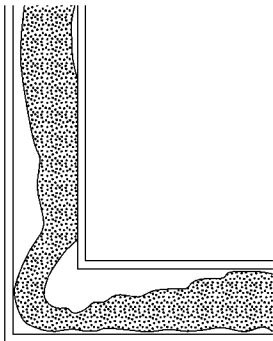


Fig. 1. Particle flow pattern near a 90° bend.

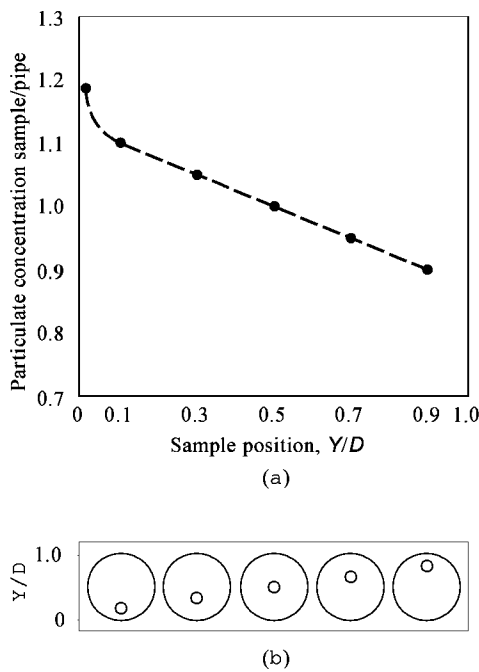


Fig. 2. (a) Particle concentration profile of liquid flowing in a pipe, where Y/D = the ratio of the distance along the diameter to the diameter (•); (b) sampling position, Y .

Although fluid systems containing particulates introduce sampling difficulties, these systems do conform to one rule of good sampling. They are in motion. For example, powders should be sampled from a moving stream rather than when at rest. Also the whole of the stream should be sampled, not just a part of it. For gas systems, whole stream sampling is usually not possible; for liquids, it can only be done from the outfall of a pipe. A third rule in sampling is that small quantities should be taken frequently rather than large quantities taken infrequently. The ideal place to sample is where the sample is well-mixed.

Chemical plants are rarely well designed in terms of sampling capabilities. Thus the rules of sampling are usually not obeyed, and most sampling involves some compromise. Decisions such as the sampling procedure to be employed, the quantity of material to be taken, and the permitted tolerance in the representativeness of the sample must be made on the basis of the use to which the subsequent analysis is to be put. These decisions depend on the analytical facilities available, whether these are manual or automatic, the skill and experience of the sampling personnel, sample and analysis correlation, and cost-benefit relationship. Finally, chemical and physical changes during sampling and subsequent handling need also to be minimized.

Careful consideration of the use to which the subsequent analysis is to be put can save time and effort in the initial sampling and possible resampling operations. For mass loading determination, the collection of particulates from hot ducts may be carried out using filters. If size distribution and classification are necessary to provide data on chemical composition as a function of size, then an in-duct impactor should be used to classify the sample directly at the temperature and location of the sampling nozzle. Agglomeration and aggregation resulting from the formation of liquid and solid bridges upon cooling can transform a flowable particulate into a solid mass which may not be easily redispersed.

Representative sampling demands a knowledge of the chemistry and chemical reactivity of the species being sampled. If a sample is being withdrawn from a hot, high pressure reactor in which the carrier gas is oxygen-free, the extracted sample, when cooled, stored, and analyzed, rarely fails to contact the atmosphere at some stage. This contact can at the least induce hydrolysis and changes in the particulate surface. If size analysis is needed, the atmospheric reaction may not affect the results, depending on the increase of agglomeration or aggregation, but it can significantly affect the surface chemical composition. Physically, the sample may be representative, but chemically it may not be. It is important that steps be taken to minimize possible physical and chemical changes before the sampling operation begins.

The quantity of sample required comprises two parts: the volume and the statistical sample size. The sample volume is selected to permit completion of all required analytical procedures. The sample size is the necessary number of samples taken from a stream to characterize the lot. Sound statistical practices are not always feasible either physically or economically in industry because of cost or accessibility. In most sampling procedures, samples are taken at different levels and locations to form a composite sample. If some prior estimate of the population mean, $\bar{x}_p$, and population standard deviation, σ , are known or may be estimated, then the difference between that mean and the mean, $\bar{x}_s$, in a sample of n items is given by the following:

$$\bar{x}_s - \bar{x}_p = \frac{t\sigma}{n^{1/2}}$$

where t is the test function for differences between mean values. Rearrangement of this equation gives the following:

$$n = \left(\frac{t\sigma}{E} \right)^2$$

where E is the maximum allowable difference between the estimate to be made from the sample and the actual value (4). For example, for a standard deviation of 0.187, the number of samples required to assure with 95% confidence that the average quality of a lot lies within ± 0.15 of the mean, the number of samples required is

$$n = \left(\frac{2 \times 0.187}{0.15} \right)^2 = 6.22 (\sim 7) \text{ samples}$$

If the standard deviation of the lot cannot be estimated, a sampling program of greater sample size is required to generate an estimate of the standard deviation for future sampling operations. In some cases, sample size can be increased and sampling costs reduced by the use of automatic samplers. These offer a substantial reduction in labor costs but an increase in capital costs.

The need for skill and experience on the part of sample designers and personnel cannot be overemphasized in chemical plant sampling. Safety precautions are of the utmost importance. Necessary steps must be taken to document the hazards involved in an operation and to ensure that the staff are well-trained, informed, protected, and capable. Except for bulk powder sampling, most chemical plant sampling is hazardous and difficult and must be designed with care. The following discussions are based on the assumptions that most of these decisions have been made and a satisfactory sampling procedure has been planned.

Gases

By far the largest proportion of gas sampling operations in industry is carried out for environmental reasons and the sampling methods employed have been thoroughly researched and are well documented (5–12). The preparation, precautions and equipment requirements involved in the sampling of air pollution sources are applicable to most other gaseous environments (see AIR POLLUTION CONTROL METHODS).

Before a source analysis program is undertaken, it is important to decide which information is really required. Sampling sites must be selected with care. Choice of the site can significantly affect accuracy and cost. Care must also be taken in the selection of sampling points at the site. Measurement usually involves the determination of temperature, concentration, and characterization of the gas contaminants. It also requires the mass rates of emission of each contaminant, therefore concentration and volumetric flow data are required.

Sampling Site. The location of a sampling site and the number of sampling points are based on the need to obtain representative data and whether the points are restricted by access problems. Sampling sites should be at least eight stack or duct diameters downstream and two diameters upstream from any disturbance. A disturbance is interpreted as a bend, expansion, contraction, valve, baffle, or visible flame. Often this type of siting is impossible and compromises have to be made. For rectangular ducts, the cross section should be divided into 12 equal rectangular segments having a sampling point at the centroid of each. For circular ducts, diameter >6.1 m (240 in.) the cross section should be divided into 12 equal areas and sampling points located at the centroid of each annulus on two perpendicular diameters (Fig. 3). For smaller-diameter ducts the number of sampling points can be reduced. If the criteria for eight diameters downstream and two diameters upstream cannot be met, extra sampling points are required (Table 2).

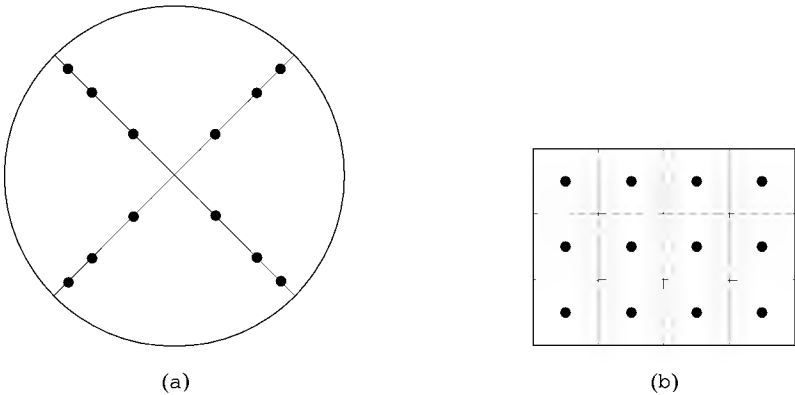


Fig. 3. Location of traverse sampling points. (a) Cross section of stack showing location of traverse points (•) on perpendicular diameters. The circular cross section is divided into three equal areas at $0.5774\,r$, $0.8165\,r$, and r , where r is the radius. Sampling points are at the centroids of these areas at $0.38\,r$, $0.70\,r$, and $0.911\,r$. (b) Cross section of rectangular stack divided into 12 equal areas having traverse points (•) at the centroid of each area.

Table 2. Number of Stack Diameters from Flow Disturbance^a

Distance from stack, number of stack diameters		Number of traverse points	dia	Diameter, m
Downstream	Upstream			
>8	>2	8		
7	1.75	12		>6.1
6	1.5	16		
5	1.25	20		
<5	<1.25	24		

^a Ref. 13.

Measurement of Gas Velocity and Temperature. Stack-gas velocity is determined at each traverse point, based on the gas density and a measurement of the average velocity head, using a pitot tube (Fig. 4) (13). The measured velocity pressure is the difference between the total pressure as measured against the gas flow and the static pressure measured perpendicular to the gas flow. The type-S pitot tube has the advantage of easy entry into a duct and a low incidence of plugging when large amounts of particulates are present. However, the tube does not give a direct measurement of velocity pressure, rather it must be calibrated for the velocity being measured. Details of this procedure have been documented by the U.S. EPA (13). Correction factors of 0.78–0.92 have been reported.

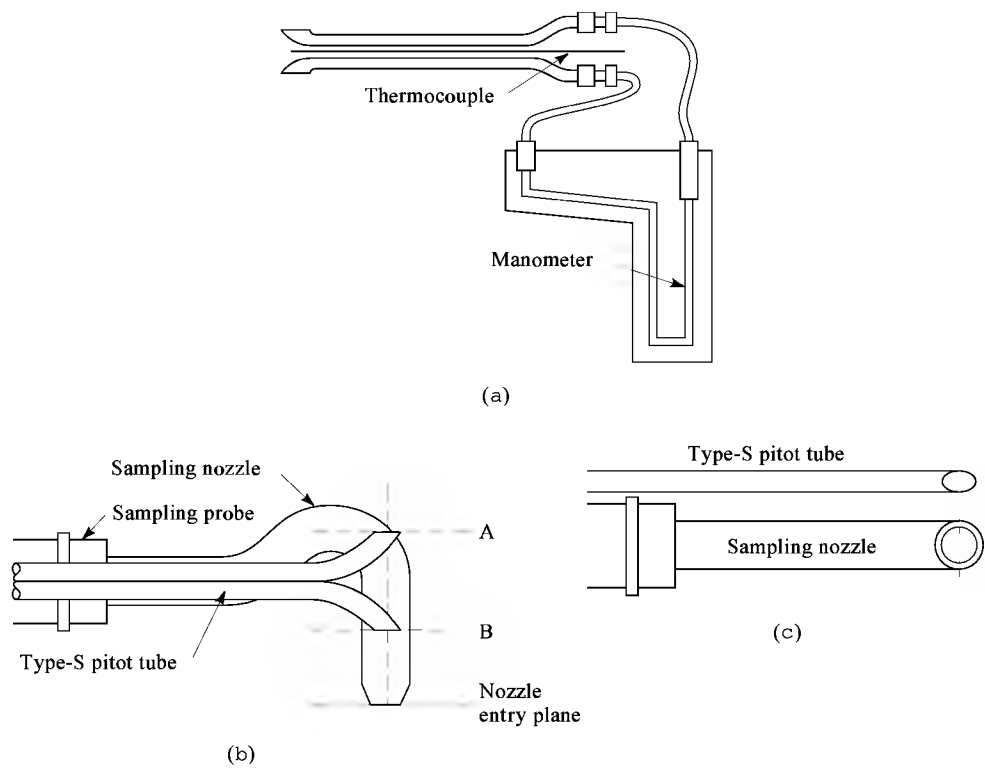


Fig. 4. (a) Type-S pitot tube and thermocouple; (b) side view of the correct pitot tube configuration when used in conjunction with a sampling nozzle, where A is the static pressure opening plane and B is the impact pressure opening plane; and (c) bottom view.

During sample and velocity traverses, the S-pitot tube is rarely used in isolation. It is necessary to measure stack or duct temperature profiles to determine variation in gas distribution, and this is usually done at the same time as the velocity profiles. If the temperature measured at each sampling or traverse point are the same, then a single gas sampling point suffices later; but if the temperature varies by more than 5%, then the sample point must be withdrawn from the traverse points selected in the initial stack survey. For most purposes, the pitot tube is combined with a thermocouple and sampling nozzle in a sampling assembly. As the presence of other components can significantly affect the correction coefficient applied to the S-type pitot tube, the placement of various components is critical to minimize aerodynamic interference. These placements are shown in Figure 4b and 4c (13).

Sample Extraction. Once the velocity and temperature profiles have been taken, gas samples can be withdrawn. In the sampling of noncondensable gases which are free of particulates, the gases are extracted from the duct by one of the following methods: a single-point grab sample, a single-point integrated sample, or a multipoint integrated sample. The last method is applicable for the collection of CO₂, CO, O₂, excess air, and nitrogen from any process in which other gases and compounds are not present in concentrations likely to affect the result. The sampling probe may be made of stainless steel, borosilicate, quartz glass, aluminum, copper, or Teflon. For a grab sample, a one-way squeeze bulb is attached to the probe to extract the gas. For integrated sampling, the train shown in Figure 5 is recommended (13). A glass or Pyrex wool filter is inserted in the probe tip to remove any unwanted particles, to prevent blockage of the train, and to prevent gas adsorption or reaction on cooling. A pump sucks the gas through a cold trap to remove moisture and then through a rotameter to measure flow rate. The gas is then collected in a gas bag where it is stored prior to analysis. Alternatively, the moisture content is measured directly by passing the gas through a set of impingers in an ice bath (Fig. 6) (13).

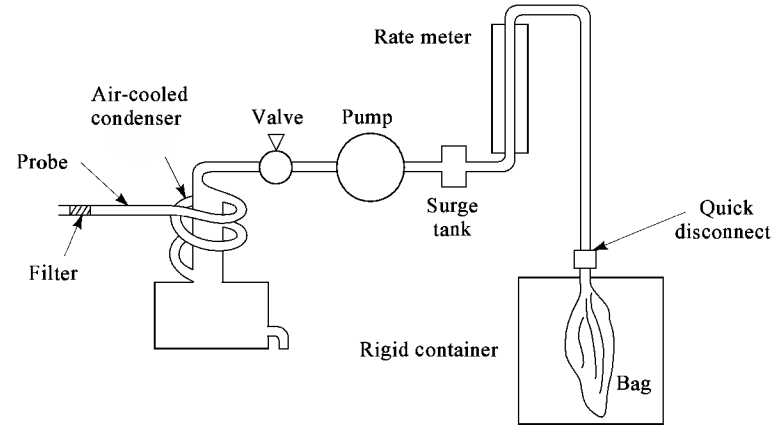


Fig. 5. Integrated gas sampling train.

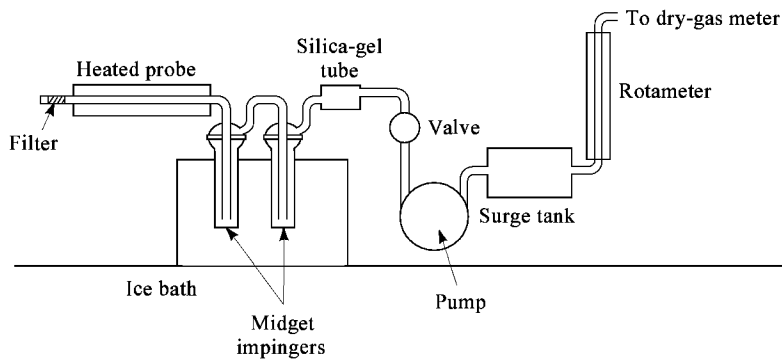


Fig. 6. Moisture sampling train.

Specific trains are recommended for other gases. Sulfur compounds (qv), eg, SO₂, SO₃, H₂S, and mercaptans, are generated during combustion, ore roasting, paper (qv) manufacturing, and other industrial operations (see SULFUR REMOVAL AND RECOVERY). In most instances, total sulfur is measured. The sampling train for SO₂ is shown in Figure 7 (13). Sulfur dioxide, which is highly reactive, is sampled through a heated or well-insulated probe and sample line to an impinger train. The probe contains a quartz or Pyrex wool filter to remove particulates. The midget bubbler in the impinger train contains 80 wt % isopropyl alcohol to remove SO₃; the first two midget impingers contain 3 wt % H₂O₂ and the third midget impinger is dry. The gas is drawn into an silica-gel drying tube. The SO₂ is estimated by titration with 0.01% N barium perchlorate using thorin as an indicator.

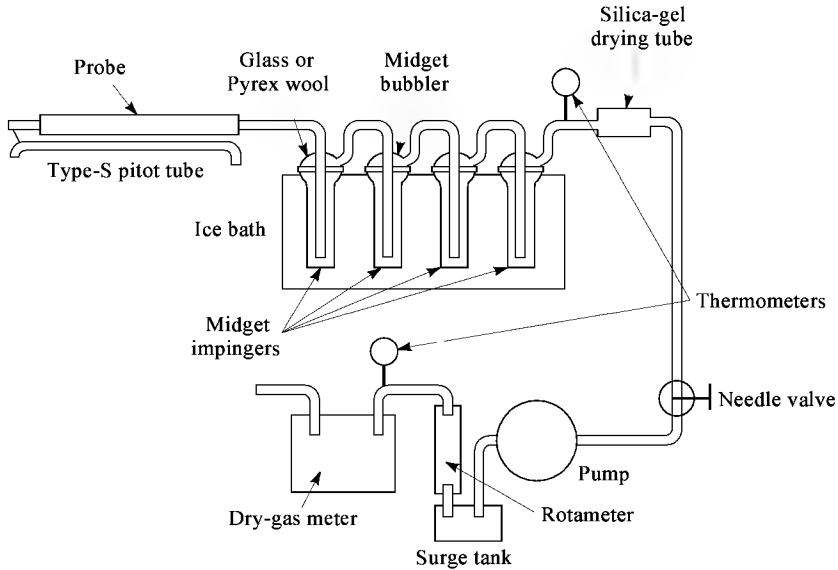


Fig. 7. SO₂ sampling train.

Nitrogen oxide sampling is simpler. This gas is drawn into an evacuated sample flask containing dilute sulfuric acid and hydrogen peroxide. The flask is shaken and allowed to stand for 16 h before the flask pressure is measured. Then the solution is made alkaline, and the nitrogen oxides are determined by the phenoldisulfonic colorimetric test.

Sample Extraction When Particulates Are Present. Different designs of probe and train may be required depending on the reason for sampling particulates. The simplest train is for the determination of mass loading only. Representative sampling of particulates is obtained using only isokinetic sampling, ie, when the velocity of the gas in the sample nozzle is the same as the gas velocity in the pipe, duct, or stack. If the sampling velocity in the sample nozzle exceeds that in the duct, the extracted sample is deficient in coarse, because large particles having high inertia do not follow the streamlines and are not collected. If the sampling velocity in the sample nozzle is less than that in the duct, the extracted sample has an excess of coarse, because small particles follow the streamlines around the duct and are not collected. The effect becomes significant for particle diameters in excess of 3–5 μm. Furthermore, the plane of the sample nozzle must be perpendicular to the gas flow or probe misalignment errors become significant.

For mass loadings, particulate matter is withdrawn isokinetically and collected on a glass fiber filter maintained at 120 ± 4°C. The appropriate sampling train is illustrated in Figure 8 (13). The sample nozzle is made of stainless steel having a sharp pointed leading edge. The taper is on the outside to provide a constant internal probe diameter. The probe is usually of a buttonhook or elbow design in order to meet alignment requirements. The first two impingers contain a known amount of water, the third is usually empty, and the last contains silica-gel. Particulate matter, present as solid or liquid at the sampling temperature, is collected on a preweighed filter and determined by weighing. Organic condensable matter is collected in the water, extracted with chloroform and then ether, and weighed after evaporation to dryness. The water is also evaporated to dryness and the residue reported as inorganic condensable matter. Particulates are sometimes present in liquid rather than solid form, eg, as acid mists (14). An alternative system is available for high volume sampling (15).

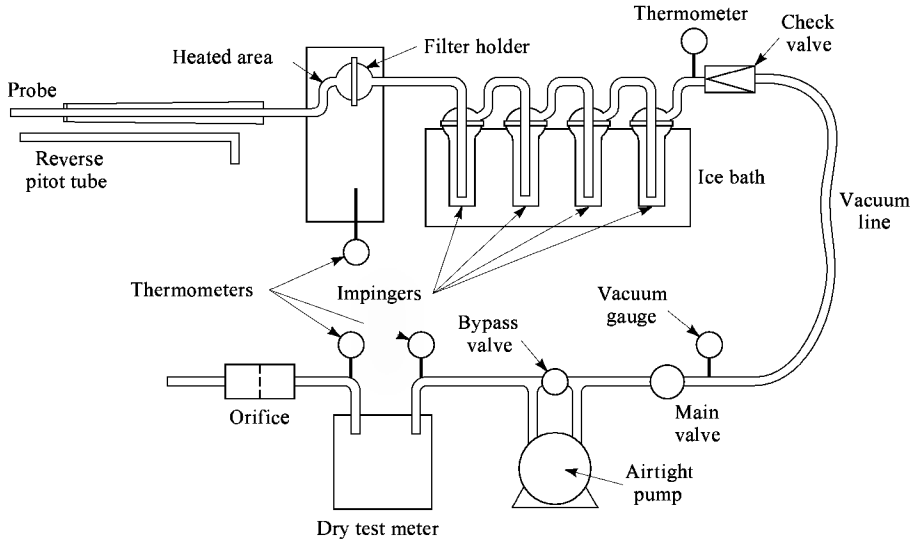


Fig. 8. Particulate sampling train.

Figure 9 shows the sampling train for sulfuric acid mist collection (13). The first impinger contains 80 wt % isopropyl alcohol and the second and third contain 3 wt % H₂O₂. The first impinger and filter retain the acid mist and SO₃; the next two retain the SO₂. After sampling, the filter is added to the contents of the first impinger and the total acid is titrated and reported as sulfuric acid.

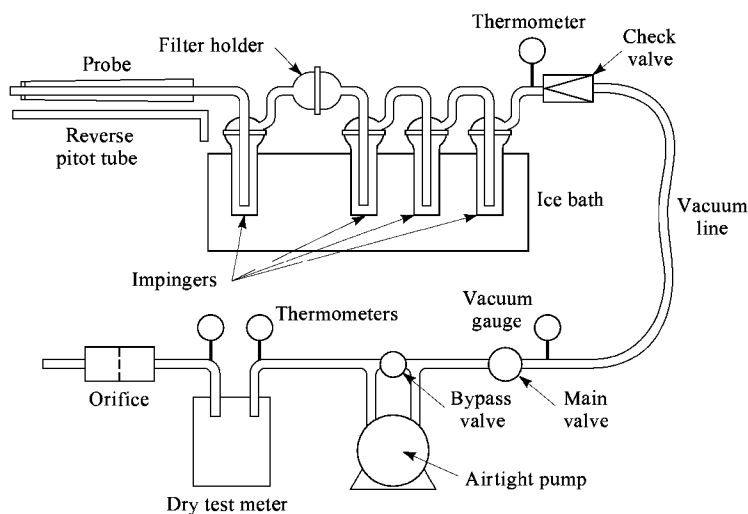


Fig. 9. Sulfuric acid mist sampling train.

In many cases, mass loadings are not sufficient. For example, size distributions are needed to correlate respirable fractions in health physics and other operations. Under conditions of high temperature followed by condensation, solidification, and cooling, redispersion of filtered particulates into their original size distribution is rarely possible. In such cases, size distributions are measured directly by use of an in-stack impactor. The Anderson stack sampler (16) is a typical example. The conditions that are necessary for mass loading measurements, ie, isokinetic sampling and nozzle alignment, are required in the use of this type of sampler. The sampler classifies the particulates drawn into the housing according to aerodynamic diameter and deposits different sizes on different collection plates. These can be weighed or chemically analyzed to permit calculation of aerodynamic diameter size distributions and determination of chemical composition as a function of size.

For acid mists, the Brink impactor is often used (Fig. 10) (17). The mist is first drawn through a cyclone to remove particles larger than 3 μm . A five-stage impactor is used to classify mist particles of diameter 0.3–3.0 μm .

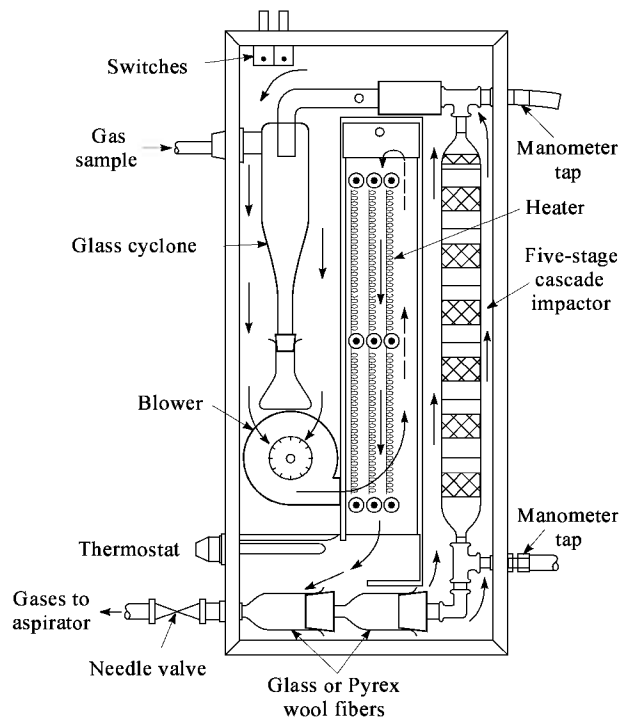


Fig. 10. Diagram of the Brink impactor.

Liquids

In the chemical industry, liquids are sampled from process vessels, tanks, tanker trucks, tank cars, ships, barges, pipelines, transfer lines, drums, carboys, cans, bottles, open lagoons, settling ponds, sewers, and open flowing streams and rivers. For simplicity, the procedures for sampling liquids are divided into three categories: tanks and similar containers, pipelines (qv), and open streams and lagoons. The problems that arise in sampling liquids containing particulates are not as critical as those that arise in gaseous systems. There are two reasons. Viscosity tends to dampen the effects of sudden changes in flow direction, and particles do not separate as readily from streamlines unless the particle masses and velocities are large.

Tanks. Sampling methods for tanks, trucks, tank cars, barges, etc, usually involve the use of fixed sample taps, thief samplers, or bottle samplers. Tanks are often equipped with stationary taps attached to pipes extending 0.6–0.9 m inside the tank. Usually three taps, one located in each third of the tank height, are sufficient. A delivery tube, long enough to reach the bottom of the sample container to allow submerged filling, is attached to the tap. Separate samples are taken from each tap, or a composite sample is taken by attaching the delivery tube to each tap in succession and filling the bottle one-third each time. When particulates are present, as in slurries or suspensions, representative samples can only be obtained from agitated tanks (see TANKS AND PRESSURE VESSELS).

A more flexible sampling system involves the use of thief tubes and bottle devices. If the tank is designed so that at least half of the cross-sectional area of the liquid surface can be exposed for sampling, samples are withdrawn in a regular grid pattern similar to that used in gas sampling. For thief sampling, the tank is divided into 12 equal areas, and the thief sampler inserted at the centroid of each area. Sample thief tubes for liquids consist of long tubes containing one or more compartments along their lengths which are isolated from the liquid by valves. At a specified depth for single-compartment tubes and upon contact with the tank bottom for multicompartment tubes, the valves are opened and the compartments filled. The valves are closed

before the thief is removed from the liquid. Each compartment is analyzed independently or several compartments may be combined to give a composite sample. The bottle sampler is a heavy-metal, perforated screen cage surrounding a one liter bottle, which can be lowered into the liquid. The metal casing weights the bottle so that it sinks. The bottle is stoppered and, after insertion, the stopper is removed by pulling on an attached line. Separate samples are taken by filling the bottle completely at different depths. In some instances, the sampler is mounted on a rigid rod fitted with a mechanical bottle opener, which allows partial filling to be carried out at a range of depths so that a composite sample may be obtained.

In many cases, access is limited to a single fixed filling hole or vent and only one port is available for thief sampling. This limits the statistical accuracy of the sample and may even give inaccurate information. The sampling thief for drums, carboys, bottles, and cans consists of a hollow tube. This is inserted to a measured depth into the container and the open end closed using a finger or thumb. The tube is removed and the withdrawn sample deposited in a container. Aspirator bulbs or peristaltic pumps are sometimes used to remove larger samples.

Pipes and Pipelines. Samples may be withdrawn both from closed pipe cross sections and from the outfall of open pipes. The simplest system for the former consists of an in-pipe sampling probe as shown in Figure 11 (18). In the chemical industry, many pipe samples are taken from organic liquids, which may be toxic, highly reactive, and flammable. The amount of sample taken should be minimized in order to reduce worker exposure and sample disposal problems. Adequate ventilation must be provided because the vapors from many liquids are more dangerous than the liquids. Splash guards are necessary when sampling corrosive and toxic liquids. For these applications in-line samplers are preferred. These trap and isolate a predetermined, precise volume of liquid from the line and deliver it to a closed container. Samplers can be installed on either side of the suction or discharge side of pumps. Typical devices include sampling plugs, multiport valves, and pneumatic samplers (19). The sampling plug is usually inserted in a bypass line as shown in Figure 12. The plug has both sample and vent connections. When open, the liquid pumped through the bypass line passes through the plug and can be returned to the line. When closed, a small constant volume of liquid is trapped in the plug and, on rotation of the tap, is drained into the sample bottle while the tap is simultaneously vented (19). If exposure to air can cause a problem in the process, the valve can be vented with nitrogen and then closed before the liquid in the bypass line is returned to the process. Mainstream sampling is usually performed with a sharp-edged probe facing directly into the flow at some preset point. In pressurized systems, sampling sites can rarely be statistically designed with regard to location and once installed are difficult to relocate. Care has to be taken when discharging plug samplers using pressurized systems in order to prevent the liquid venting through the valve connections (19). Sometimes a safety valve is inserted in the bypass to prevent blockage of the pump discharge.

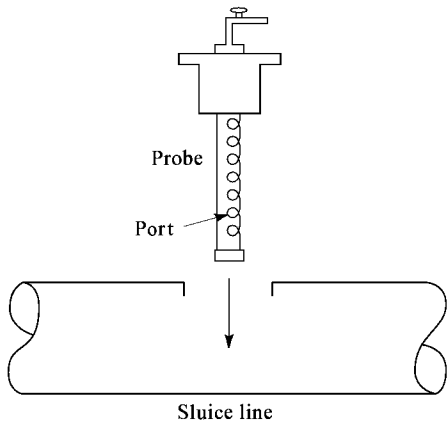


Fig. 11. In-pipe sampling probe having 0.635 cm dia sampling ports.

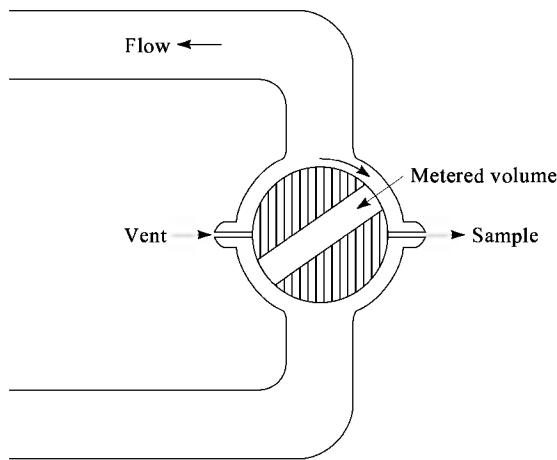


Fig. 12. Metered sampling plug.

Larger sample sizes can be obtained using multiport valves (19). During normal operation, the liquid flows through the valves in a bypass line. Often two valves are employed. A sample is extracted by turning both valves simultaneously, first to isolate the sample and then to let it discharge into the sample bottle. This is best achieved by gear-and-linkage devices, which mechanically link the two valves. Failure to do this simultaneously can result in the full line pressure being vented through the sample line (19). Thus this type of sampler presents potential hazards when used for high pressure systems.

Air-electric samplers can be installed directly in the pipe wall. One type of liquid sampler is operated by a solenoid valve that activates an air cylinder. A shaft is moved in and out of the pipe by this cylinder and samples are expelled into a container below the sampler. Sample volumes of from 2–30 mL are possible.

For low pressure pipelines that have ports open to the atmosphere, eg, sewers or closed effluent culverts, samplers are designed to sample through manholes. In a typical system, the liquid is lifted through a suction line into the sampling chamber under vacuum. When filled, the vacuum shuts off, and the sample drains into a sample jar. A secondary float prevents any liquid from reaching the vacuum pump. The suction line then drains by gravity back to the source.

A more permanent installation is provided by a chain-driven sampler, widely used in paper (qv) and steel (qv) mills, manufactured as the E Sampler by QCEC (20). A cup, which is attached to a chain positioned perpendicular to flow, travels down through the liquid flow and returns to the upper sprocket, where the sample is drained into a container. Flow-proportional timers can be installed to change the rate of sampling with flow rate (see *Flow Measurement*).

Cutter-type samplers can be installed for low pressure pipelines and enclosed troughs. These samplers contain a movable cutter connected to a flexible hose through which the sample is extracted. Such devices meet all three rules for good sampling: they sample a moving stream, the sample is made

up of many increments, and the whole stream is sampled uniformly.

For representative slurry sampling the smallest aperture of the sampler must be greater than three times the diameter of the largest particle present in the slurry. Sample overflow should not occur. The cutter is located in a housing directly in the pipeline and is driven across the whole stream by a motor-driven traverse. Two designs are shown in Figure 13. Cutter samplers can also be placed at the oufalls of pipes or weir discharges (21). Alternatively, a rotary cutter, eg, a Vezin sampler, can be used for flow rates up to 3.8 L s⁻¹ (18).

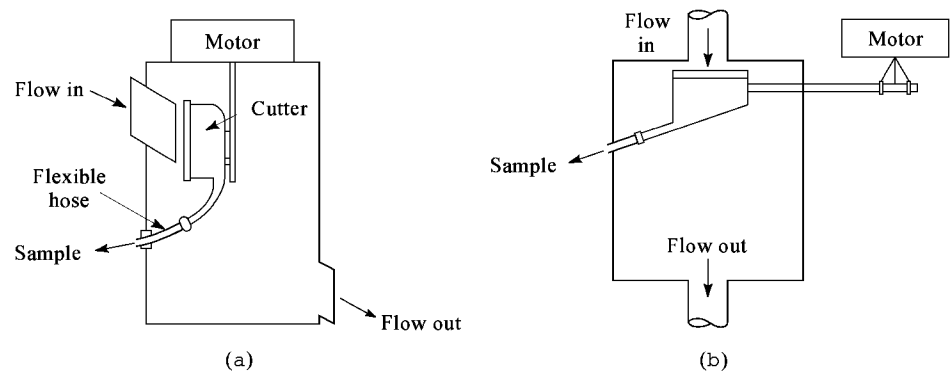


Fig. 13. Slurry samplers: (a) sloping trough cutter and (b) vertical pipe cutter.

Open Streams and Lagoons. Open-stream discharges are encountered in wastewater plants in the chemical industry (see WASTES, INDUSTRIAL; WATER POLLUTION). Settling ponds and lagoons are part of wastewater treatment plants. Details on the monitoring of industrial wastewater are spelled out in EPA regulations (22). Surface samples may be withdrawn using a dipper sampler of nonreactive material (23,24). Water-pollution abatement programs are based on information obtained by sampling (22). Representative samples are a principal concern of any abatement program. The composition and magnitude of waste flows varies widely. Flow often differs significantly on a cyclical basis. General maintenance is often performed when the plant is not in full production, usually on weekends. This can result in unusually high suspended solids loading in the discharge stream over short periods of time. Boilers may be cleaned at such times and this can give rise to slugs of very highly alkaline water. Lagoon sampling should therefore be carried out every day on a 24-h schedule unless the plant discharge is known to be free of such variations. Several samples should be taken during a 24-h period. The number depends on the variability of the discharge. Ideally the sample should be taken from a place where the flow is well mixed, eg, an oufall from well-mixed tanks, but EPA regulations are often based on the condition of the final oufall to a stream or a river, where the composition can vary widely. Consequently, the flow rate must be measured (22) so that the total waste being discharged can be calculated, and samples must be taken at high frequency in proportion to the flow. Samples are often composited over an 8-h shift or some other suitable time frame.

Many final discharges are mixed effluents. These may contain oil and particulates that tend to separate near the surface or the pipe bottom, respectively. Furthermore, the conditions promoting solids suspension, eg, turbulence, also entrains air into the sample, which can change the dissolved oxygen content significantly affecting chemical oxygen demand (COD) and biochemical oxygen demand (BOD) measurements. Plant-induced turbulence prior to discharge is allowable, but sampler-induced turbulence, eg, vigorously shaking a grab sample prior to pouring it into a sample bottle to suspend the particles, is not good practice. Procedures are also needed to avoid contamination and confusion in labeling. Reference 22 provides detailed instructions for the correct sampling methods.

Preserving the sample is more critical in wastewater sampling than in gas or powder sampling. Biochemical and chemical changes can take place fairly rapidly and thus samples must be analyzed as quickly as possible. For unstable samples, the use of composite samples is not good practice. Instead, special grab samples should be taken and presented to the laboratory for analysis within one hour of sampling. Table 3 gives recommended storage and preservation conditions for various analytical parameters.

Table 3. Recommended Storage Procedures of Samples^a

Analysis	Sample storage	
	Preservation	Maximum holding period, d
total solids	cool to 4°C	7
suspended solids	cool to 4°C	7
volatile suspended solids	cool to 4°C	7
chemical oxygen demand	H ₂ SO ₄ to pH<2	28
biochemical oxygen demand	lag develops, must use fresh sewage seed	2 ^b
acidity-alkalinity	refrigeration at 4°C	14
biochemical oxygen demand	refrigeration at 4°C	2
calcium	H ₂ SO ₄ to pH<2	180
chemical oxygen demand	H ₂ SO ₄ , 2 mL L sample	28
chloride	none	28
color	refrigeration at 4°C	2
cyanide	NaOH to pH>12, cool to 4°C	14
dissolved oxygen		no holding
fluoride	none	28
hardness	H ₂ SO ₄ to pH<2	180
metals		
total	HNO ₃ to pH<2	180
dissolved	filtrate:HNO ₃ to pH<2	180
nitrogen		
ammonia	H ₂ SO ₄ to pH<2, cool to 4°C	28
Kjeldahl	cool to 4°C	28
nitrate-nitrite	H ₂ SO ₄ to pH<2, cool to 4°C	28
oil and grease	H ₂ SO ₄ to pH<2, cool to 4°C	28
organic carbon	H ₂ SO ₄ to pH<2, cool to 4°C	28
pH		no holding
phenolics	H ₂ SO ₄ to pH<2, cool to 4°C	28

phosphorus	H ₂ SO ₄ to pH<2, cool to 4°C	28
solids	cool to 4°C	7
specific conductance	cool to 4°C	7
sulfate	refrigeration at 4°C	28
sulfide	zinc acetate 2 mL L sample, NaOH to pH>9	7
threshold odor	refrigeration at 4°C	7
turbidity	cool to 4°C	2

^a Refs. 25 and 26.
^b In composite sampling system.

Either grab or composite sampling may be used for lagoons. Grab samples are good for continuous flows, but cannot be used for samples proportional to the flow. They are only reliable when the discharge composition is constant. They are essential to pinpoint the times of isolated high pH or high solids content. Composite samples can be taken proportional to flow or time, but such samples provide averages of the conditions of isolated high values of pH and suspended matter to the point where the values of these factors may not be obvious from the final analysis. Another consideration is whether to sample manually or automatically. Manual sampling involves high labor but low capital costs and is necessary when the sampling position is variable. Automatic sampling can involve very high capital costs but is beneficial for high frequency and around-the-clock sampling programs. Automatic sampling can be linked with particle size measurement to give size distributions minute-by-minute. It can also be linked with robot stations for physical and chemical monitoring. Alarms can also be installed to warn of any unusual deviations from normality.

Bacteriological sampling is performed by manual techniques because of stringent sterilization requirements. Samples are taken in wide-mouthed, sterile, glass-stoppered bottles that are wrapped in paper prior to sterilization in an autoclave at 138 kPa (20 psi) or in an oven at 170°C. The bottle is unwrapped and the lower portion is held in the hand. The sample is taken with the bottle mouth in the direction of the flow. The stopper must be protected from contamination, the bottle only partially filled, and the sample stored at 4°C after sampling. For bacteriological samples withdrawn from a tap, the water should run for five minutes and then be shut off; the tap should then be sterilized by flaming before a sample is taken.

Where free chlorine is present, eg, in drinking water, it is measured on-site, and a crystal (eg, 10 mg 40 mL) of sodium thiosulfate is added to the bottle prior to sterilization to convert free chlorine to chloride.

Radioactive samples require other, special techniques. Some are discussed in Reference 22 (see RADIOACTIVE TRACERS).

Equipment. Manual sampling is performed with one-liter wide-mouthed bottles. A long-handled wide-mouthed scoop is often used for less accessible sampling points. Weighted bottles or specially designed samplers that open at any required depth under water are also used. Hand-operated pumps are used for less accessible sampling locations. A wide range of automatic samplers is commercially available (27,28). No single sampler is suitable for all sampling needs. Over 40 manufacturers supply automatic sampling equipment and these tend to give widely differing data. The sampling method is more site-dependent than any design attribute (27).

A nonproportional sampler is suitable for near-constant flow conditions. The sample is simply drawn from the waste stream at a constant flow rate. Sampling lines should be as short as possible and free from sharp bends, which can lead to particle deposition. Proportional samplers are designed to collect either definite volumes at irregular time intervals or variable volumes at equal time intervals. Both types depend on flow rate. Examples of some of these are the vacuum and chain-driven wastewater samplers. Other types, which have cups mounted on motor driven wheels, vacuum suction samplers, and peristaltic pump samplers, are also available (26,27).

Samplers must be designed and constructed to withstand the chemical composition extremes present in the individual discharges. Corrosion-resistant fabrication must be used in the equipment that comes in contact with many chemical industry discharges (see CORROSION AND CORROSION CONTROL).

Solids

Solids occur in several forms in the chemical industry. Raw material from natural deposits are compacted in the ground and sampling is performed during the exploration stages. This type of material is typified by minerals and fossil fuels. Before use, these must be crushed, ground into particulate form, mixed, cleaned, and stored. Material may be stored on the ground or in a silo, bin, or hopper. Products in particulate form are usually stored in ships, barges, railcars, drums, boxes, cans, bags, etc. During manufacture these are transported by conveyors, pipes, and chutes and are packaged with the use of free-flowing streams, pneumatic conveyors, spouts, and gravity chutes. Sampling may be required before, during, or immediately after any one of these operations, and different sampling methods are used for most of them (see CONVEYING).

Natural Deposits. Natural deposits, eg, minerals and fossil fuels, are located by drilling operations. An auger, eg, a screw or worm, is turned in the earth and pulled out, and material is scraped from the auger for analysis. Alternatively, samples can be taken by hollow core drills which, when withdrawn, enclose a core of the earth that is representative of the strata through which the drill has passed. Such core samples are used in geological surveys for fossil fuels. As the drill drives deeper into the strata, each core is extracted and placed in a shallow box and coded so that a complete cross section of the geological strata can be reconstructed. From this, the relative thickness of coal and mineral seams can be directly measured.

Segregation. Free-flowing powders have a natural tendency to segregate. During transportation, fine particles percolate through coarse ones to give vertical segregation. During pouring, into a heap or into a container, coarse particles travel farther than fine ones leaving, for example, for a centrally poured heap, an excess of fine particles at the center of the heap. Segregation depends therefore on the previous history of the powder. Even nonflowing powders may segregate during manufacture or handling and, once segregated, remain segregated unless subjected to a mixing operation (see POWDERS, HANDLING).

Whenever possible a powder should be sampled when in motion; and the whole of the stream of powder should be taken for many short increments of time in preference to part of the stream being taken for the whole of the time. Observance of these rules coupled with an understanding of the manner in which segregation takes place leads to the best sampling procedure. Care and skill in abstracting samples is needed and cannot be overemphasized.

Sampling Stored Material. A very large number of possible systems of stored material exists from which a gross sample has to be extracted. The method to use depends primarily on whether the powder is stationary or moving, and whether it is cohesive or free flowing. It is usual to assume that the powder was mixed before storage. If this assumption is invalid, then the homogeneity of the powder depends on its history. Thus, a nonflowing material that has been segregated prior to storage remains segregated. For a free-flowing material, segregation can occur during transfer from the mixer to the storage container.

Stored Nonflowing Materials. Nonflowing materials are composed of very fine cohesive powders, sticky materials, moist material, or fibrous solids. These may be stored in small containers such as drums or bags, or in large containers such as a truck or railcar. In order to obtain a representative sample from a small container, it is preferable that the material be premixed or all the material be passed through a sampling device such as a spinning riffler or Vezin sampler. For large containers, samples may be obtained from the surface or from the body of the material. Surface sampling is usually carried out using a scoop because of simplicity. An assumption is that the powder at the sampling point is representative of the bulk, ie, the powder was mixed before storage. Accuracy is increased by taking more than one sample. Samples should be analyzed separately and combined in later analyses if the variation between samples is at an acceptable level. Body sampling is usually done manually or by a power-driven thief. For the latter, a split-tube thief, ie, a tube having a slot running its entire length, is used with a sharp cutting edge at the lower end of the tube. The thief is inserted in the center of the container and rotated to cut out a core of material. The thief is withdrawn and the material is scraped from it for analysis.

Stored Free-Flowing Material. It is practically impossible to representatively sample stationary free-flowing powder because of the severe segregation that has almost certainly occurred. If there is no alternative but to sample this material, several samples should be taken and analyzed separately, so that an estimate can be made of the reliability of the measured parameter. For free-flowing materials stored in small hoppers, drums, cans, boxes, and

bags, static sampling by the use of thief samplers is quite common. The typical sampler consists of two tubes, one fitting snugly inside the other. The tip of the outer tube is sharply pointed, and holes are cut in both tubes to mate at a specific angle of rotation. The outer tube is rotated until the holes are closed, and the thief is inserted firmly into the powder bed. At the desired depth the inner tube is rotated until the holes mate and the thief is opened. Powder flows into the inner tube cavities, after which the tube is again rotated to the closed position. The thief is withdrawn and re-opened to expose the samples. Inner tubes are compartmentalized to ensure that depth profiling can be conducted. Alternatively, a composite sample can be formed.

For larger storage bins, automatic auger samplers are used. A solenoid-controlled air cylinder opens and closes an aperture in the sampling tube, and the sample is drawn to the discharge point by a motor-driven auger. Depth profiles can also be conducted by scoop sampling while the container is being emptied. Sampling thieves have a tendency to arch and thus are vibrated in order to break the arch and allow powder to enter the holes. This is not a recommended practice, however, because fines are more readily admitted. The recommended procedure is to get the powder in motion and obtain a sample using a cutter or Vezin-type sampler.

In the chemical industry it is common to sample small heaps by coning and quartering. The heap is first flattened at the top and then separated into four equal segments with a sharp-edged board or shovel. The segments are drawn apart and frequently two opposite quadrants are recombined and the operation repeated until a small enough sample has been generated. This practice is based on the assumption that the heap is symmetrical. This is rarely so, however, and the withdrawn sample is usually nonrepresentative. This method is no more accurate than the scoop or thief sampling methods, which are simpler to carry out. Coning and quartering should never be used with free-flowing powders, because the most important segregation property is particle size. When poured into a heap, the fines tend to percolate to the center whereas the coarse particles roll down to the outside.

Flowing Streams. All free-flowing powders are transported at some time during manufacture as flowing streams. Hoppers are emptied by screw conveyors. Solids are transported to bagging operations by pneumatic conveyors, and most solids pass through transfer points in gravity-flow pipes and chutes. Even small samples in boxes, bags, cans, bottles, etc, can be made to flow by emptying them into volumetric feeders. Sampling is carried out on the resultant stream.

Sampling from pneumatic conveyors parallels gas sampling. The exception is that solids loadings can be as high as 50 kg of solids per kg of gas. Commercially available samplers extract particles directly from a transport line. Fixed position samplers are mounted directly on the pneumatic conveyor pipe. Devices are available which extract samples from the product stream by the projection of a sample tube into the flow. Particles impact on the tube and fill the open cavity. The tube is then withdrawn, and an internal screw discharges the collected material (20). In another model, the RX Sampler (manufactured by Gustafson) (29), samples are withdrawn using compressed air.

Sampling from screw or drag conveyors is effected using slide-gate samplers (20,29). Sampling probes cannot be mounted directly into the lines, therefore sliding gates are positioned on the bottom of drag housing or on the bottom or sides of screw conveyors. An air cylinder opens and closes the discharge gate, permitting powder to fall through the discharge tube into a sample container.

When a sample is to be collected from a conveyor belt, the best position for collecting the increments is where the material falls in a stream from the end of the belt. If access at such a point is not possible the sample must be collected from the belt. The whole of the powder on a short length of the belt should be collected. The particles at the edge of the belt, however, may not be the same as those at the center; particles at the top of the bed may not be the same as those at the bottom. If the belt can be stopped, the sample may be collected by inserting a frame into the stream consisting of two parallel plates shaped to fit the belt. The whole of the material between the plates is then swept out. A scoop can be used to obtain an increment, but this operation can be hazardous if the belt is moving. Bristol Engineering Company (30) manufactures a belt conveyor system that has only one moving part in contact with the material. An arm sweeps across the belt to remove the sample.

Sampling from a continuous stream may be continuous or intermittent. In continuous sampling a portion of the flowing stream is split off and frequently further divided subsequently. In intermittent sampling the whole stream is taken for many short increments of time at fixed time intervals. These increments are usually compounded and samples for analysis taken from this gross sample. Consignment sampling is carried out on a single consignment, eg, a truck or wagon load.

For large tonnages, samples taken from conveyors can represent large quantities of material which need to be further reduced. Often, a traversing cutter is used as a primary sampler, and the extracted sample is further cut into a convenient quantity by a secondary sampling device. Secondary traversing-type samplers are marketed by the Denver Equipment Corporation, QCEC, and Gustafson, Inc. This equipment is satisfactory for many applications. Limitations which restrict use, however, include the following: although comparatively easy to design and build in a new plant, it is frequently difficult and expensive to retrofit an existing plant, primarily owing to space requirements; the quantity of sample obtained is proportional to product flow rate and this can be inconvenient when the plant flow rate is subject to wide variations. On the other hand, where a plant's daily average is required, this is a necessary condition; and it is difficult to enclose the sampler to the extent required to prevent the escape of dust and fume when handling a dusty product.

Commercial samplers are available that combine a traversing-type sampler and an unacceptable table sampler. An alternative design is the radial cutter or Vezin sampler. These samplers vary in size from a 15-cm laboratory unit to a 152-cm commercial unit.

Efficiency. Sampling of bulk solids from grinding circuits or chemical plants represents tons of material per day. A primary sampler generally removes 10–100 kg as a gross sample, which is then subdivided into 1–10 kg laboratory samples by a secondary device. The samples may be examined as taken or may be crushed prior to examination. Measurement samples are needed by the laboratory in gram or milligram quantities and on the basis of subsequent measurement decisions are made as to the quality of tonnages of bulk material. It is therefore essential that the measurement sample be representative of the bulk. Bias at any stage of the sampling affects the final result (31).

In one sampling method study, sugar, coarse sand, and fine sand were used to test several laboratory sampling procedures (32), eg, cone and quartering; scoop sampling, which consists of plunging a scoop into the powder and removing a sample; chute splitting; table sampling; and rotary riffling. The rotary sample divider consists of a hopper and vibratory feeder, from which powder is made to flow in a constant stream. The powder falls onto a rotating circular tray subdivided into various compartments. Several versions of this instrument are available handling from 40 liters down to a few grams.

Scoop sampling is particularly prone to error because the whole of the sample does not pass through the sampling device, and the sample is taken from the surface where it may not be representative of the mass. In an attempt to achieve a good mix the sample container is shaken prior to sampling. In a sampling table the material is fed to the top of an inclined plane in which there is a series of holes. Prisms placed in the path of the stream break the stream into fractions. Some powder falls through the holds and is discarded. The powder remaining on the plane passes on to the next row of holes and prisms. More is removed, and so on. The powder reaching the bottom of the plane is the sample. The chute splitter consists of a V-shaped trough along the bottom of which is a series of chutes alternately feeding two trays placed on either side of the trough. None of these methods gives as efficient a sample as the rotary sample divider. The results of this study are given in Table 4. Rotary sampling is by far the best analytical sampling method to use for solids. This method follows all the rules for good sampling.

Table 4. Reliability of Solids Sampling Methods

Method	Standard deviation, %	Estimated sample error, % ^a
cone and quartering	5.76	19.2
scoop sampling	6.31	21.0
table sampling	2.11	7.0
chute riffling	1.10	3.7
rotary riffling	0.27	0.9
random variation	0.09	0.3

^a Values are maximum.

BIBLIOGRAPHY

"Sampling" in *ECT* 1st ed., Vol. 12, pp. 84–91, by H. W. Eckweiler, U.S. Bureau of Customs, and pp. 91–95, by C. L. Dunn, Hercules, Inc.; in *ECT* 2nd ed., Vol. 17, pp. 744–762, by C. A. Bicking, The Carborundum Co.; in *ECT* 3rd ed., Vol. 20, pp. 525–548, by R. Davies, E. I. du Pont de Nemours & Co., Inc.

1. *ASTM Standards Index*, American Society for Testing Materials, Philadelphia, Pa., 1993.
2. T. Allen, *Particle Size Measurement*, 4th ed., Chapman and Hall Methuen, London New York, 1990.
3. J. H. Rushten and J. Hillestad, paper presented at *The 24th Midyear Meeting of the American Petroleum Institute*, Preprint No. 52, May 1964.
4. ASTM E122-5, *Standards Designation*, American Society for Testing Materials, Philadelphia, Pa., 1993.
5. BS 3405, *Simplified Methods for Measurement of Grit and Dust Emission from Chimneys*, British Standards, London, 1961.
6. BS 893, *Methods of Testing Dust-Extraction Plant and the Emission of Solids from Chimneys*, British Standards, London, 1940.
7. C. J. Stairmand, *Chem. Eng.* **29**, 31 (1951).
8. E. B. Sansone, *Am. Ind. Hyg. Assoc. J.*, 487 (Sept.–Oct. 1993); *Sampling Airborne Solids in Ducts Following a 90° Bend*, Ph.D. dissertation, University of Michigan, Ann Arbor, 1967.
9. *Determining Dust Concentration in a Gas Stream*, Power test Code No. 27, American Society of Mechanical Engineers, 1957.
10. E. A. Wolfe, *Gas Flow Rate and Particulate Matter Determination of Gaseous Effluents*, Bay Area Air Pollution Control District 1480, San Francisco, Calif., 1961.
11. *Source Testing Manual*, No. 434, Los Angeles Air Pollution Control District, Los Angeles, Calif., 1963.
12. H. J. Paulus and R. W. Thron, in A. C. Stern, ed., *Stack Sampling in Air Pollution*, 3rd ed., Academic Press, Inc., New York, 1976, p. 3.
13. U.S. EPA *Regulations on Standards of Performance for New Stationary Sources*, 40 CFR 60, Appendix A, Reference Methods, Washington, D.C., 1993.
14. ASTM D3685-92, *Standard Test Method for Sampling and Determination of Particulate Matter in Stack Gases*, American Society for Testing Materials, Philadelphia, Pa., 1992.
15. ASTM D4536-91, *Standard Test Method for High Volume Sampling for Solid Particulate Matter and Determination of Particulate E*, American Society for Testing Materials, Philadelphia, Pa., 1991.
16. *Anderson Stack Sampler*, Anderson 2000 Inc., Atlanta, Ga.
17. J. A. Brink, *Ind. Eng. Chem.* **50**, 645 (1958).
18. ASTM D5013-93, *Sampling Waste from Pipes and Other Point Discharges*, American Society for Testing Materials, Philadelphia, Pa., 1993.
19. B. G. Lovelace, *Chem. Eng. Proc.* **51** (Nov. 1979).
20. Technical data, Quality Control Equipment Co., Des Moines, Iowa.
21. D. K. Fields, *Soc. Mining Eng.*, 1486 (Nov. 1979).
22. *Handbook for Monitoring Industrial Wastewater*, U.S. EPA, Washington, D.C., Aug. 1973.
23. ASTM 5358-93, *Standard Practice for Sampling with a Dipper or Pond Sampler*, American Society for Testing Materials, Philadelphia, Pa., 1993.
24. *Samplers and Sample Procedure for Hazardous Waste Streams*, EPA-600 2-80-018, U.S. EPA, Washington, D.C.
25. "Environmental Protection Agency, Rules and Regulations," *Fed. Reg.* **49**(209) (Oct. 26, 1984).
26. *Standard Methods for the Examination of Water and Wastewater*, 18th ed., National Environmental Research Center, Washington, D.C., 1992.
27. *An Assessment of Automatic Sewer Flow Samplers*, PB-25,987, U.S. Dept. of Commerce, Washington, D.C., 1976.
28. *A Survey of Commercially Available Automatic Wastewater Samplers*, EPA 600 4-76-05, U.S. EPA, Washington, D.C.
29. Technical data, Gustafson, Inc., Dallas, Tex.
30. *DS 3*, Sampler Bulletin 101, Bristol Engineering Co., Yorkville, Ill.
31. B. H. Kaye, Ph.D. dissertation, London University, 1961.
32. T. Allen and A. A. Khan, *Chem. Eng.* **238**, CE108–CE112 (1970).

Reg Davies
E. I. du Pont de Nemours & Co., Inc.

SAPONINS.

See STEROIDS.

SCANDIUM.

See LANTHANIDES.

SCALEUP.

See PILOT PLANTS; PROCESS CONTROL.

SCREEN PRINTING.

See DYES, APPLICATION AND EVALUATION; ELECTROPHOTOGRAPHY; PRINTING PROCESSES; THERMOGRAPHY.

SCREENS, SCREENING.

See SIZE ENLARGEMENT; SIZE MEASUREMENT OF PARTICLES; SIZE REDUCTION.

SCRUBBERS.

See ABSORPTION; ADSORPTION; COAL CONVERSION PROCESSES, CLEANING AND DESULFURIZATION; GAS, NATURAL.

SEALANTS

A sealant is a material that is installed into a gap or joint to prevent water, wind, dirt, or other contaminants from passing through the joint or gap. This joint or gap may be a fixed joint, but is often an expansion joint which may also be called a working joint. Sealants, which can also be defined by how they are tested, are rated by their ability to stretch, twist, bend, and be compressed while maintaining their bulk properties so that they do not tear apart under stress. A most important rating of a sealant in many applications is the movement ability of the sealant. The adhesion required of a sealant is simply the strength to hold the sealant in position as it is stressed and strained.

Adhesives are used to transfer loads and are typically designed with much higher tensile and shear strengths than sealants. The most important rating of an adhesive in many applications is the determination of how much load it can handle. Some sealants are used as adhesives and some adhesives as sealants and thus arises the occasional blurring of their roles. If the material's primary function is the exclusion of wind, water, dirt, etc, it is a sealant.

Performance Characteristics

Movement Capabilities. The movement capability of a sealant is the amount of displacement the sealant can endure in extension or compression without failing. One standard test used to rate movement capability is ASTM C719, Standard Test Method for Adhesion and Cohesion of Elastomeric Joint Sealants Under Cyclic Movement. This test subjects a sealant to cyclic movement of 3.2 mm h under alternating conditions of high and low temperatures as well as water immersion. The movement capability is typically reported at $\pm x\%$, where the positive value refers to extension and the negative value to compression. A sealant with $\pm 50 \pm 25\%$ rating can be repeatedly extended by 50% or compressed to 25% of the original joint width after initial cure. In general, sealants with lower modulus of elasticity can handle higher movements.

Sealants can be broadly divided into classes according to the amount of movement they can successfully handle. High performance sealants such as silicones, urethanes, and polysulfides can typically handle movements of 25% or higher. Medium performance sealants such as some acrylics can handle movements of 10–25%. Low performance sealants such as butyls, putties, and caulks accommodate movements under 10%. Movement ratings are commonly reported by sealant manufacturers, but these ratings are dependent on the substrate, cure time, rate of movement, etc. Any manufacturer's literature must be interpreted with care, as some manufacturers' literature, for instance, may only test their sealant dry, at room temperature, and claim a large movement ability. If an application in question has a working joint that will be exposed to high and low temperature as well as water, the sealant to be specified should be tested to the full rigors of the ASTM C719 test. Such testing should be done by an independent laboratory validated by an independent validating organization.

Movement ratings and tests such as ASTM C719 are useful in that they expose the sealant to conditions close to actual field conditions. However, these tests are typically carried out on freshly cured sealants, cured without any stress. In actual installations, sealants are installed in working joints and cure while joint movement occurs. More recent studies have evaluated the performance of sealants under movement during cure (1–3). Results indicate that it is possible to have a premature sealant failure on account of movement during cure. This failure may be one of several possible modes. One of the more serious modes of failure is the cracking or splitting of the sealant surface. This seems to be related to the length of time needed for a sealant surface to develop an elastic skin; sealants that develop elastic surfaces rapidly seem to resist cracking better. However, even with one-part sealants that form elastic surfaces rapidly, continuous movements of greater than 15% during cure can cause failures. If large movements during cure are expected, two-part sealants

that cure uniformly in depth may provide better resistance to failure because two-part sealants generally cure faster than corresponding one-part sealants. Another technique for reducing large joint movement on outdoor applications in hot weather is to seal during early evening hours. This allows approximately 12 hours of cure to take place before significant joint movement occurs.

Modulus. The measure of the stress of a sealant at a specific strain is referred to as the modulus of elasticity, sometimes called the secant modulus. This important sealant property describes the force exerted by a sealant as it is stressed. Because a primary function of sealants is to adhere to the substrates it is in contact with, the forces generated by a joint opening or closing are transmitted by the sealant to the substrate–sealant bond line. For this reason, it is important to know the modulus of the sealant and also the strength of the substrate. The use of a high modulus sealant on a weak substrate, such as asphalt, concrete, or sandstone, can result in stresses higher than those which the substrate can tolerate, and in a higher potential for failure, often through cohesive failure of the substrate. Normally, concrete is thought of as a high strength substrate, which it is in compression. However, by comparison it is quite weak in tension. Sometimes, upon closer inspection, perceived adhesive failures are actually cohesive failure of the substrate near the substrate–sealant interface caused by such a modulus mismatch.

Although sealant manufacturer's literature commonly reports modulus values, these values must be interpreted carefully. Specimen sizes, test rate, cure conditions, and the time a sealant has been allowed to cure when tested can all have a significant effect on modulus. Therefore, for a true comparison, sealants should be evaluated by a standard test that examines all sealants by the same procedure. In general, the longer a sealant has been allowed to cure, the more realistic the modulus data.

Durability. A primary factor in sealant durability is its ability to resist decay from environmental elements. For most typical applications this includes extremes of high and low temperature, water, oxidation, and sunlight.

Weatherability. One of the more destructive elements is exposure to sunlight; specifically, ultraviolet (uv) light. All sealants are affected by weathering, but there is much difference in the effect of weathering on different sealants. Most silicones are stable with respect to uv exposure. Urethanes and polysulfides show effects of uv exposure, but can be formulated with uv absorbers to provide reasonable lifetimes in most applications. However, there are exceptions in all classes of sealants and specifiers must be careful to look for test data that has proven a specific sealant's durability. The source of the test data is also important; data from an independent testing laboratory is generally apt to be more reliable.

A sealant's location on a building as well as the building's geographical location can greatly impact the sealant's durability. In the northern hemisphere, a sealant installed on the north side of a building has much less exposure to uv and high temperatures as compared to a sealant installed on the south side of a building, especially if the substrate is dark-colored. This sheltered exposure can extend the useful life expectancy of sealants that are prone to degradation in outdoor weather. Conversely, sealants installed in southern states, such as Florida, Arizona, Texas, and California, and on the south side of a building have a high degree of exposure to harmful uv and high temperatures as compared to any of the northern states. Such a location can accelerate the decay of a sealant prone to uv degradation.

A sealant's durability is often estimated by accelerated testing that exposes the material to high temperatures, high humidity (condensation), and uv light. Because there is much debate on how to estimate actual lifetime from accelerated testing, the duration of accelerated test exposure varies with the test method and may range from hundreds to thousands of test hours. One study of weathering effects on the surface of sealants indicates that 2000 hours of accelerated exposure is roughly equivalent to two years of full sunlight exposure in Florida with each of several different generic sealant types tested (4). Another study of the correlation of accelerated weathering to sealant modulus changes suggests eight weeks of accelerated weathering roughly corresponds to one year of weathering at Garston, England (5). In contrast, most official ASTM weathering tests for sealants in the 1990s require only 250 or 500 hours in an accelerated weathering machine. Therefore, data cited by manufacturers as having passed ASTM accelerated weathering tests may have little meaning in the real world. For example, two weathering tests for highway sealants, ASTM D3406 and D3408, require only 160 hours in an accelerated weathering machine. In actual use, sealants in these applications receive full sun exposure and much longer times in weathering machines are needed to obtain meaningful data. In general, silicones provide the best resistance to decay from water, temperature extremes, and uv light because the polymer is inherently stable. However, some urethanes and polysulfides can also successfully seal joints trouble-free for many years by incorporating sufficient uv and oxidation stabilizers. A key factor in a successful, long-life seal is for the specifier to determine the degree of exposure to harmful elements the sealant is expected to experience in its lifetime and demand test data that approximate that lifetime.

Adhesion Life. A second key factor in determining the durability of a sealant is the ability of the sealant to adhere to the substrate through its lifetime. A sealant may have excellent resistance to uv effects, but if it has poor adhesion performance and fails adhesively, it is of little use. The same can be said of a sealant with superior adhesion characteristics but poor resistance to uv. Either situation results in a short performance life.

A sealant's adhesion is commonly studied by 180 degree peel tests such as ASTM C794 or by tensile adhesion joints tests such as ASTM C719. The adhesion test protocol should simulate actual field conditions as closely as possible. Sealants often have good adhesion to dry substrates, but this adhesion may be quickly destroyed by water. Because most sealants are exposed to water over their lifetime, adhesion testing should include exposure to water for some length of time. ASTM C719 is one of the better tests to determine a sealant's adhesion durability as it exposes sealants to seven days of water immersion.

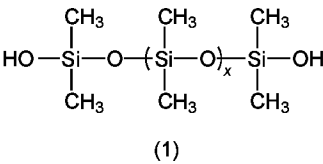
The modulus of elasticity can also influence the adhesion lifetime. Some sealants may harden with age as a result of plasticizer loss or continued cross-linking. As a sealant hardens, the modulus increases and more stress is placed on the substrate–sealant adhesive bond. If modulus forces become too high, the bond may fail adhesively or the substrate may fail cohesively, such as in concrete or asphalt. In either case the result is a failed joint that will leak.

Another way to determine durability is to find successfully sealed, existing field installations. Sealant manufacturers often have case histories of successful installations. Adhesion performance can vary widely with sealant type, substrate type, and cleaning method. For this reason, it is important to understand the sealant's sensitivity to cleaning practices. Often with difficult-to-adhere substrates, a primer is used. Sealant manufacturers can provide recommendations as to which substrates require primers, what type of primers should be used, and how they should be applied.

Sealant Types and Formulations

Silicones. Commercially available silicone sealants are typically one of three curing types: moisture-reactive (curing) sealants, moisture-releasing (latex) sealants, and addition-curing sealants. Of these three types, moisture-curing silicones make up the vast majority of silicone sealants sold.

Moisture-Curing Silicones. The formulation of moisture-curing silicones includes a silicone polymer, filler, a moisture-reactive cross-linker, and sometimes a catalyst. The most common silicone polymer used in sealant formulations is an alternating silicon–oxygen backbone with methyl groups attached to the silicon such as the silicone polymer (1).



Specialty silicone polymers with trifluoropropyl groups replacing half of the methyl groups are used to make solvent-resistant materials. Diphenyl substitution for the methyl groups on the polymer can lower the brittle point of a silicone sealant from –65 to –110°C (6). Viscosity for silicone polymers can vary widely from less than 100 mPa·s (cP) to over 100,000 mPa·s (cP). Most cure through a reactive end group. The end group can be one of several different functionalities, but is usually hydroxyl. The saturated backbone and high Si–O bond energy provides inherent stability to sunlight. Because of this stability, silicones, unlike other sealants, need no special uv absorbers. Silicone polymers also exhibit little change in physical properties with temperature as compared to organic systems. This results in sealants that are as easy to extrude in cold weather as in warm weather. It also means that the cured sealant shows less change in properties over temperature as compared to other sealants. Another fundamental property of silicone polymers is the high vapor

transmission rate. Because most one-part silicone sealants cure by reaction with water vapor, this high transmission rate allows for fast cure in depth as compared with one-part organic-based sealants.

Silicone polymer plasticizers have historically been used in many formulations. These plasticizers (qv) are of the same Si–O backbone as the functional polymers but generally are terminated with trimethyl groups which are unreactive to the cure system. This nonreactivity means that, if improperly used, the plasticizer can migrate from the sealant and stain certain substrates. Staining has been a widely publicized flaw of silicone sealants, but the potential of a formulation to stain a substrate can be minimized or eliminated with proper formulation work. In general, this is accomplished by not using plasticizers for formulations developed for stain-sensitive substrates.

Silicone polymers when cured into elastomers by themselves are weak, gel-like materials. For this reason, fillers must be used to provide reinforcement. The type of fillers (qv) used in silicone sealants varies widely; two of the most common fillers are fumed silica and calcium carbonate. Fumed silica, a highly reinforcing filler, is usually added in amounts ranging from 6 to 20^o ɔ. Silica is most often used when a high strength sealant is desired. Several silicas having different surface areas are available and surface treatment with silanes may be used as well.

For lower modulus sealants, either ground or precipitated calcium carbonates are commonly used. These fillers are much less reinforcing than silica and may take up 50^o ɔ or more of a formulation. The surface of calcium carbonate is often treated to place a stearate group on the filler. By using various filler types, loadings, and filler treatments, the rheological nature of the sealant can vary from nonsag to self-leveling with low or high extrusion rates, and tensile strengths can vary from 0.35 to 7 MPa (50–1000 psi).

The moisture-reactive cross-linkers used in silicones are of the form R_nSi(OR')_{4-n}, where n = 0 or 1 and R may be any organic group, such as methyl, ethyl, or vinyl. R' also varies; acetoxy, alkoxy, oxime, and propenoxy are among the most typical. Common cross-linkers are listed in Table 1, and a typical formulation for a one-part silicone sealant is given in Table 2.

Table 1. Common One-Part Silicone Cross-Linkers and Their Leaving Groups

Cure system	Cross-linker	CAS Registry Number	Leaving group
alkoxy	CH ₃ Si(OCH ₃) ₃	[1185-55-3]	methanol
acetoxy	CH ₃ Si(OOCCH ₃) ₃	[4253-34-3]	acetic acid
oxime	CH₃ CH₃ Si(ON=C)₃ CH₂CH₃	[22984-54-9]	methylethylketoxime
propenoxy	CH₂=CH—Si(OC=CH₂)₃ CH₃	[15332-99-7]	acetone

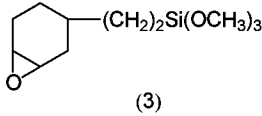
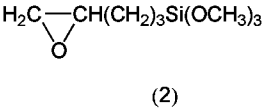
Table 2. Typical One-Part Silicone Sealant Formulation

Component	High modulus sealant, %		Low modulus sealant, %	
silicone polymer	70–85		40–60	
silicone plasticizer	0–5		0–20	
fumed silica	6–12			
calcium carbonate			40–60	
cross-linker	3–8		3–8	
silane adhesion promoter(s)	0–1		0–1	
catalyst (if needed)	<1.5		<1.5	

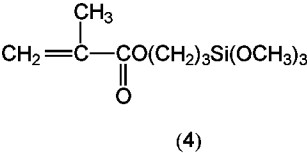
Some of the silicone cross-linkers are reactive enough on their own to cure a silicone sealant without the use of a catalyst. Most, however, require a catalyst, usually a tin carboxylate or an organotitanate.

The cure rate of a silicone sealant is dependent on the reactivity of the cross-linker, catalyst type, catalyst level, the diffusion of moisture into the sealant, and the diffusion of the leaving group out of the sealant. For one-part sealants, moisture diffusion is the controlling step and causes a cured skin to form on the exposed sealant surface and progress inward. The diffusion of moisture is highly dependent on the temperature and relative humidity conditions.

Most moisture-curing silicones have good general adhesion to a variety of substrates. However, adhesion can be markedly improved with different combinations of silanes. The more common silane adhesion promoters are categorized as amine functional, eg, 3-aminopropyltrimethoxysilane [13822-56-5], 3-aminopropyltriethoxysilane [919-30-2], or N-(2-aminoethyl)-3-aminopropyltrimethoxysilane [1760-24-3]; as epoxy functional, eg, 3-glycidoxypropyltrimethoxysilane [2530-83-8] (2) and 2-3(3,4-epoxycyclohexyl)ethyltrimethoxysilane [3388-04-3] (3); as mercapto functional, eg, 3-mercaptopropyltrimethoxysilane [4420-74-0]; or as methacrylate functional, eg, 3-methacryloxypropyltrimethoxysilane [2530-85-0] (4).



(3)



(4)

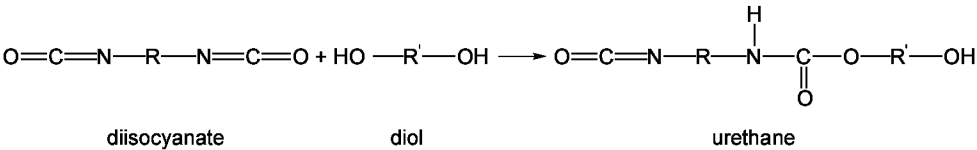
Each of these silanes consist of two functionalities: the moisture-reactive functionality (usually alkoxy), and the organic functionality. It is believed that this

dual functionality allows the adhesion promoter to interact with both the substrate and the sealant. These silanes are also effective adhesion promoters for other nonsilicone-based sealants. Principal producers of these silanes are Dow Corning, H&ls, and OSI.

Moisture-Releasing (Latex) Silicones. A newer class of silicone sealants are known as the silicone latex sealants. These sealants are silicone-in-water emulsions that cure by evaporation of the emulsifying water. They should not be confused with siliconized acrylic sealants, which contain less than 2% silicone and have all the characteristics of acrylic sealants. The silicone latex polymer is prepared by first emulsifying a low molecular weight silicone polymer in water and then polymerizing it to the desired molecular weight. Inherent to emulsion polymerization is the ability to produce high molecular weight polymers at a low emulsion viscosity. Next, a silicone cross-linker is added with a condensation catalyst. The cross-linker, the structure of which is similar to those described previously, must diffuse through the water phase and into the siloxane phase where it can react with the silicone polymer. The water pH must be carefully controlled to prevent the cross-linker from reacting with water before it reaches the silicone phase. Fillers and freeze–thaw additives are also typically added to silicone latex sealants (7).

Addition-Curing Silicones. Addition-curing silicones in general are two-part systems that cure by the platinum-catalyzed reaction of a silicone hydride with typically a vinyl group attached to silicon. Because no by-products are generated by the cure, there are few volatiles and no shrink in thick sections. The cure can take place at room temperature or be accelerated by heat. Addition-curing silicones have little inherent adhesion and proprietary adhesion promoters or primers are often used to give adequate adhesion to substrates. The catalyst level used in these systems ranges from 5 to 500 parts per million platinum. Because this level is so small, these systems are easily poisoned by things that react with or complex with platinum.

Urethanes. The basis for urethane chemistry is the reaction of an isocyanate group with a component containing an active hydrogen.



The first step in formulating a urethane sealant is to prepare what is commonly called the prepolymer, typically by reaction of a hydroxy-terminated polyether with a stoichiometric amount of a diisocyanate. Each hydroxy group reacts with one end of every diisocyanate molecule.

Although polyethers are the main building blocks of the prepolymer, other materials such as polyesters, polythioethers, and polybutadienes are also used. The choice of the starting polymer provides certain characteristics to the resulting urethane sealant; especially important are polymer size and branching. Most urethanes use blends of polymers to achieve desired properties. With proper additives, polyether-based urethanes provide good general weathering properties and abrasion resistance. Typical urethane sealants have a high temperature limit of 135°C, and a maximum use temperature of 80°C. Polyester-based sealants have good fuel and weather resistance, but poor low temperature flexibility and poor resistance to water. In contrast, polybutadiene-based urethanes have good resistance to water, but have poor weather resistance and a limited temperature range. Polythioethers provide chemical resistance in elastomeric materials that cannot be achieved in polyether urethanes (8).

Polyethers are available in a wide range of molecular weights, from ca 400 to greater than 6000. Diols and triols are most common, but higher functionalities are also available. Common diisocyanates employed are toluenediisocyanate [1321-38-6] (TDI) or 4,4'-methylenediphenyldiisocyanate [101-68-8] (MDI) with toxicity issues leading toward more use of MDI. Aromatic diisocyanates are most common, but aliphatic diisocyanates are also used. Although more expensive than aromatic isocyanates, aliphatic isocyanates have better color and uv stability. They are used in applications where added stability is needed and where price can support the extra cost. The other unreacted isocyanate end of the prepolymer is still reactive to active hydrogens and is the basis for two-part urethanes. The first part consists of the urethane prepolymer and the second part contains a polyether. By mixing the two parts together, reactions occur between the isocyanate prepolymer and the polyether, resulting in a network having elastomeric properties. Most urethane sealants are of these two- or even three-component types that are mixed on-site.

One-part urethane sealants (Table 3) are more complicated to formulate on account of an undesirable side reaction between the prepolymer's isocyanate end and water vapor which generates carbon dioxide. If this occurs, the sealant may develop voids or bubbles. One way to avoid this reaction is to block the isocyanate end with phenol and use a diketamine to initiate cure. Once exposed to moisture, the diketamine forms a diamine and a ketone. The diamine reacts with the isocyanate end on the prepolymer, creating a cross-link (10). Other blocking agents, such as ethyl malonate, are also used (11). Catalysts commonly used in urethane formulations are tin carboxylates and bismuth salts. Mercury salt catalysts were popular in early formulations, but have been replaced by tin and bismuth compounds.

Table 3. Typical One-Part Urethane Sealant Formulation^a

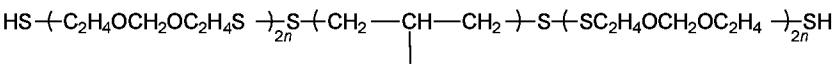
Ingredient	Wt %
diisocyanate prepolymer	30–35
filler	30–45
plasticizer	15–35
uv absorber	1–3
antioxidant	1–2
dehydrating agent	1–3
adhesion promoters	1–3
thixotrope	2–3
solvent	3–5
coloring paste	2–3

^a Ref. 9.

The fillers used in urethane formulations are similar to those used in silicones; calcium carbonate, talc, clays, and silica are among the most common. On account of the undesirable reaction of isocyanates with water, fillers used in urethane formulations must be dry.

Urethane sealants have good inherent adhesion to most substrates, but silane adhesion promoters are often used to improve this adhesion. Epoxy-, amino-, and mercapto-functional silanes are the most common because of their dual reactive nature. The silane end can react with surface hydroxyls; the epoxy, amino, or mercapto end can react with the isocyanate.

Polysulfides. Polysulfide sealants were the first high performance synthetic elastomeric sealants produced in the United States. The basic polymers are mercaptan-terminated (HS–R–SH), with molecular weights ranging from 1000 to ca 8000. Cross-link concentrations range from 0.5 to 2.0 mol %. Curing occurs through the mercaptan groups by oxidation and results in an S–S linkage. In addition to the terminal SH groups, pendent SH groups are located along the polymer chain and contribute to cross-linking. A typical polysulfide polymer may be represented as follows:



The value of *n* is approximately 6 for Morton International Polymer LP-3, 24 for LP-32.

Polysulfide polymers provide inherent resistance to fuel and quite good resistance to alkali. In contrast to the silicone polymers, they have low gas

permeability. Although low gas permeability is an asset in certain applications such as single-seal insulating glass, it also causes slow cure for one-part moisture-activated sealants where the moisture must diffuse through the sealant to initiate cure. Because of the slow cure of the one-part sealant, the most common form of polysulfides is the two-part mix-on-site sealants.

Two of the more recently developed polysulfide polymers are the mercaptan-terminated polyoxypropylene urethane polymer and the polythioether polymer. The urethane-backbone-based polymer is used in many sealant formulations for insulating glass applications. The thioether backbone contains sulfur, but no S–S bonds, which are the weakest part of the conventional polysulfide polymer. This polymer improves the thermal stability and reduces the gas–liquid permeability.

Early formulations typically used lead dioxide as the oxidative curing agent. More recently, curing agents such as manganese dioxide and sodium perborate monohydrate have been used (Table 4). Different curing agents have different effects on the properties of the polysulfide sealant. Therefore, the choice of curing agent depends on the sealant application. Early polysulfides used polychlorinated biphenyls (PCBs) as plasticizers, but these have been completely replaced with phthalate esters or other plasticizers (qv).

Table 4. Typical One-Part Manganese Dioxide-Cured Polysulfide Formulation^a

Ingredient	Wt %
polysulfide polymer(s)	30
calcium carbonate (coated)	30
phthalate plasticizer	25
titanium dioxide	8
molecular sieve	0.5
adhesion promoter	1.5
fumed silica	0.5
manganese dioxide (active)	2.0
barium oxide paste (75% active)	1.5
stearic acid	0.2

^a Ref. 12.

Common fillers used in polysulfide formulations are ground calcium carbonate, precipitated calcium carbonate, carbon blacks, and clays. The pH of the ingredients used in the formulation can impact the cure rate either positively or negatively. Organic amines are used to increase the cure rate, and acids to decrease the cure rate. Small-particle-size calcium carbonate is often coated with organic treating agents, such as stearic acid, to improve the dispersing properties. However, such an acid treating agent can slow the cure rate.

Adhesion promoting silanes are often added to improve adhesion to various substrates. As is the case with urethane sealants, silanes with a dual-reactive nature are typically used. Examples of such silanes are mercapto- and epoxy-functional silanes. Organic titanates may also be used.

Acrylics. There are two principal classes of acrylic sealants: latex acrylics and solvent-release acrylics.

Latex Acrylics. High molecular weight latex acrylic polymers are prepared by emulsion polymerization of alkyl esters of acrylic acid. Monomer, water, surfactants, and an initiator are mixed and polymerized until the acrylic monomer is depleted. Two types of monomers are used to vary polymer properties. High *T_g* monomers such as methyl methacrylate and vinyl chloride improve durability and hydrophobicity, whereas polar-functional monomers such as hydroxyethyl acrylate are used to improve adhesion. The maximum level of solids for the latex polymer is approximately 60%. In typical formulations (Tables 5 and 6), above this point the viscosity increases rapidly and the emulsion stability is poor. In relatively low solids (high water) content formulations, rather severe shrinkage occurs during cure. This can introduce stress and may be one of the reasons most latex acrylics are of lower performance and lower movement ability. The surfactants used are of special concern to sealant formulations because they can interfere with adhesion if improperly used. One approach to solve this problem is to incorporate the surfactant into the polymer backbone during polymerization. This approach, which places the surfactant in an ideal location to stabilize the emulsion, does not allow the surfactant to migrate through the aqueous phase and interfere with adhesion because the surfactant is connected to the backbone (13). The emulsion polymers are compounded into sealants by adding fillers, plasticizers, freeze–thaw stabilizers, thickeners, and adhesion promoters. As is true of the silicone latex sealants, the acrylic latex sealants are easy to apply and clean with water.

Table 5. One-Part Pigmented Siliconized Acrylic Latex Sealant^a

Ingredient	Wt %
acrylic latex polymer (polymer and water)	30
calcium carbonate	56
plasticizer	8
mineral spirits	1.7
propylene glycol	1
titanium dioxide	1
ammonium hydroxide (14%)	<1
preservative package	<1
nonionic surfactant	<1
inorganic dispersants	<1
organic dispersants	<1
defoamer	<1
associative thickener	<1
silane adhesion promoter	<1

^a Ref. 13.

Table 6. One-Part Clear Acrylic Latex Sealant Formulation^a

Ingredient	Wt %
acrylic latex polymer (polymer and water)	78.7
plasticizer	11.2
fumed silica	2.2
nonionic surfactant	1.7
amino silane	0.8
ammonium hydroxide	0.7

other	4.7
^a Ref. 14.	

Solvent-Releasing Acrylics. Another class of acrylic sealants are the solvent-releasing acrylics. Acrylic monomers are polymerized in a solvent. The molecular weight of the polymer is lower than in the latex acrylics because of the inherently higher viscosity of the medium. However, the percentage of solids is approximately 80% vs the 60% common to latex acrylics. The natural adhesion of most of the solvent-releasing acrylics produces some of the best unprimed adhesion in the sealant industry. However, slow, continual cure generally produces large compression sets and limits their use to low movement applications. Also, the relatively high amounts of solvent and traces of acrylic monomer in these formulations limits their use to outdoor applications, usually in construction.

Butyls. Butyl-based materials are sold in the form of preformed tapes, thermoplastic hot melts, and one-part solvent-releasing sealants. Butyl polymers are made by the copolymerization of 97–98 mol % isobutylene with 2–3% isoprene. Another butyl-based polymer, polyisobutylene, is produced by the polymerization of isobutylene. Both polymers are available in a wide range of viscosities. These polymers are low cost and have the lowest moisture vapor transmission rates of the common sealant types. Formulations of butyl-based sealants also include plasticizer, filler, and tackifier resins. Polybutenes are common plasticizers for butyl sealants. They are made from the polymerization of a mainly isobutylene feedstock and are available in several different viscosities. Butyl sealants are often filled with carbon black, but other fillers such as talc and calcium carbonate are also used, especially for lighter-colored sealants. Tackifier resins, such as the pentaerythritol ester of rosin, are added to provide tack or adhesion to various substrates. Solvents, such as mineral spirits, are used for the one-part solvent-releasing formulations (Table 7). As the solvent leaves the typical one-part butyl, the sealant hardens and loses its elastomeric ability. This limits the use of solvents to low movement applications where durability is not of high concern. The hot melts, on the other hand, are quite plastic in nature and cracks tend to self-heal. Hot melts usually have good longevity in special applications such as the primary seal in insulating glass (Table 8).

Table 7. One-Part Solvent-Releasing Butyl Sealant Formulation^a

Ingredient	Wt %
butyl polymer	22
talc	31
calcium carbonate	21
polybutene plasticizer	10
mineral spirits	6
tackifier (terpene phenolic resin)	4
rutile titanium dioxide (pigment and uv protection)	3
thixotrope (fumed silica)	2
other	<1

^a Ref. 15.

Table 8. Hot-Melt Polyisobutylene Formulation^a

Ingredient	Wt %
polyisobutylene polymer	25
resin	30
amorphous polypropylene resin	20
low density polyethylene	15
carbon black	10

^a Ref. 15.

Manufacture and Processing

Almost all sealants contain a mixture of a powdered filler incorporated into a viscous liquid, which results in a viscous liquid sealant having a paste-like consistency. Batch mixers such as Ross Mixers and other planetary types are common to the industry. Also, continuous compounding of sealants can be accomplished using compounders such as the Werner-Pfleiderer (WP) Compounder. Mixers can be broadly classified into low and high shear mixers (see MIXING AND BLENDING). In general, low shear mixers are able to disperse low surface area fillers, such as ground calcium carbonate (average particle size of 1 μm), into the sealant polymers quite well. However, high surface area fillers, such as the high surface area fumed silicas (surface area >200 m²/g), usually require higher shear rate mixers, such as the WP Compounder or sigma-blade batch mixers. High surface area fillers incorporated into sealants on a low shear mixer may result in material having a grainy appearance as the mixer is not able to disperse the small particles of filler sufficiently.

Processing conditions can have a dramatic effect on sealant rheology, cure time, and physical properties. Typical processing variables are mixer speed (rpm), time, temperature, and vacuum. Order of ingredient addition is also important. For one-part moisture-curing silicones, it is important to keep the raw materials dry and to minimize exposure of the finished sealant to moist air. Failure to do so can result in the formation of small cured gels in the sealant. Dry raw materials are also critical when manufacturing one-part urethanes and polysulfides. The isocyanate end of the prepolymer is very reactive with moisture and can lead to gassing from carbon dioxide formation. The catalysts used to cure one-part polysulfides are moisture-sensitive and must be protected from moisture during manufacture. Three methods used to dry urethane and polysulfide raw materials are azeotropic distillation with toluene, vacuum drying, and addition of a desiccant to the formulation. Stoichiometry of the urethane prepolymers is critical for achieving the desired sealant properties. Specifically, this means closely controlling the ratio of isocyanate to hydroxyl.

Temperature also plays an important role. The ordered reactivity of the diisocyanate may change if the temperature is too high, resulting in both ends of the diisocyanate reacting with the polyether's hydroxyl groups. Latex sealant compounding has a special requirement of its own. The processing equipment must achieve specific particle size, specific particle size distribution, and specific molecular size within the particle. High shear equipment is usually required to produce stable emulsions. Incorrect processing conditions can produce a variety of bad effects, including the ability to break a smooth emulsion, which results in a two-phase mixture of water and large curds.

Economic Aspects

Since the introduction of the first polysulfide sealants in the 1950s, the sealant market had grown by 1992 into a \$1.4 billion industry. Sealant use is generally divided into four categories: transportation, construction, consumer, and industrial. Table 9 lists the market segment value of each use as of 1992 and Table 10 the market share of the various material types. Continued growth is expected, especially for silicones, urethanes, and acrylics. Table 11 lists suppliers of sealants.

Table 9. 1992 U.S. Sealant Consumption by End Use Application ^a

Market segment	Value, \$ × 10 ⁶
transportation ^b	520
construction ^c	415
consumer	350
industrial	140
Total	1425

^a Ref. 16.
^b Includes PVC plastisols and automobile windshield materials.
^c Includes commercial and residential building, heavy construction, maintenance, and insulating glass.

Table 10. Market Share and Projected Growth Rate ^a

Product type	1992 Value, \$ × 10 ⁶	1992–1997 Annual growth rate, %
silicone	325	3–6
butyl rubbers	215	0
acrylics	165	1–2
polyurethanes	145	3–6
polysulfides	90	1 to 3
other	485	0
Average		1–2

^a Ref. 16.

Table 11. Suppliers of Sealants

Sealant	Supplier
silicones	Dow Corning, General Electric, Wacker Silicones, Rhone-Poulenc, Bayer, Shin-Etsu
polyurethanes	Essex Speciality Products, Mameco International, Pecora Chemical, Courtaulds Aerospace Inc., Sika, Tremco, ChemRex
polysulfides	Morton International, ChemRex, Pecora, Courtaulds Aerospace Inc., Chemseal
acrylics	DAP, 3M, Tremco, Rohm and Haas (polymers)
butyls (hot melt)	H. B. Fuller, Tremco, Exxon (polymers)

Specifications

Numerous specifications exist for sealants. One of the more important specifications for construction sealants is ASTM C920. This specification is based on experience with high performance sealants in the structural sealing and glazing applications. Among the most confusing specifications are the highway pavement sealant specifications. Each U.S. state Department of Transportation (DOT) lists specifications for pavement sealants, almost all of which are less stringent than ASTM C920; only a few are more rigorous. The confusing issue is that different chemical types of highway sealants have different performance specifications for the same application. For military applications, many military specifications apply and must be adhered to. In industrial applications, typically each industry has developed its own specifications to meet a specific application. For example, automobile manufacturers have a series of specifications relating to the use of sealants in cars and trucks. Aircraft sealants are another example of specifications developed for an application. An excellent reference for the numerous specifications is available (17). In addition to specifications, other testing codes may be important, depending on the application. For fire or electrical applications, testing approved by the Underwriters Laboratory (UL) is necessary. For sealants that are used in food processing, U.S. FDA testing is needed.

Health and Safety Factors

Sealant Manufacturing. Most sealants use mineral-based fillers which may contain small amounts of crystalline silica. If crystalline silica is present, dust control is important to prevent inhalation of these particles. Crystalline silica is a known cause of silicosis, a debilitating disease of the lung. Another common safety concern in sealant manufacturing is the use of flammable materials. Not all sealants use flammable ingredients, but for those that do, proper inerting and grounding are needed to prevent potential explosions.

For silicone manufacturing, handling of cross-linkers and catalysts in 100% pure form requires knowledge of any unusual hazard. These materials may be skin irritants, flammable, or capable of reacting with moisture to release one of the leaving groups in Table 1. Urethane sealant manufacturing requires precautions during the production of the prepolymers. Some isocyanates used to make the prepolymer are known skin irritants and can cause dermal sensitization by direct contact. Inhalation of the vapors of some isocyanates can sensitize the pulmonary tissues of certain people and may cause serious asthmatic symptoms (9). Acrylic monomers used to make emulsion polymers for acrylic sealants have varying degrees of toxicity and may require special safety and handling procedures. Also, some surfactants and freeze–thaw stabilizers may be skin irritants.

Sealant Application. The moisture-reacting silicones cure by reaction of the cross-linker with atmospheric moisture and release of the leaving groups listed in Table 1. In general, these sealants should be used in areas that have good ventilation. The leaving groups are released in small amounts, but they can accumulate in a confined or unventilated area. Acetoxy-curing silicones release small amounts of acetic acid. Oxime-curing silicones release methyl ethyl ketoxime that can be a potential skin sensitizer. Although relatively small amounts of these materials are released, it is a good idea in general to avoid skin contact. Latex acrylic sealants may be minor skin irritants on account of the surfactants from the polymerization step. Urethane sealants may also be skin irritants; skin contact should therefore be avoided. Polysulfide sealants have a characteristic odor, probably resulting from small amounts of low molecular weight, sulfur-containing components. Although the odor is somewhat objectionable, the polymers are not known to be skin or eye irritants. Primers and cleaners historically have been mainly neutral organic solvent-based, but many are being formulated to minimize or eliminate solvent. However, a careful review of the material safety data sheet is fundamental before use of any sealant.

Uses

Each class of sealants has certain attributes inherent to the polymer on which it is based. These attributes often define the sealant's applications and limitations. Table 12 summarizes each type of sealant's key attributes, principal markets, and limitations.

Table 12. Summary of Sealants

Sealants	Key attribute	Principal market	Limitation
silicone	excellent uv stability, fast cure on account of high moisture vapor transmission, good high tempera-ture stability, good low temperature flexibility	structural glazing, weatherproofing, insulating glass (secondary seal), pavement, con-sumer, premium housing	some not abrasion-resistant, some not recommended for below-grade or continuous water im-mersion, generally not paintable, some pick up dirt
urethane	tough, abrasion-resistant, good adhesion to many substrates, good chemical resistance	weatherproofing, automotive, park-ing structures, insulating glass	80°C (175°F) tempera-ture limit, limited uv stability, most not used for typical glazing, some have limited stability in water, not suitable for structural glazing
polysulfide	excellent chemical resistance, low gas permeability	insulating glass, aircraft sealants, fuel-resistant applications, below-grade applications	slow cure for one-part sealants, limited uv stability, not suitable for structural glazing
acrylic	good uv stability, easy to apply and clean (latex), low cost, good adhesion	residential housing, consumer	low movement capability, limited low tempera-ture flexibility
butyl	low gas permeability, low cost, paintable	insulating glass (primary seal), automotive, low cost housing	low movement capability, limited durability

Future developments are likely to feature the production of more silicone sealants that do not pick up dirt, more latex acrylic sealants that have high performance properties, urethanes that have improved uv stability, and high performance polysulfides that are made in the United States.

BIBLIOGRAPHY

"Sealants" in *ECT* 3rd ed., Vol. 20, pp. 549–558, by R. B. Seymour, University of Southern Mississippi.

1. J. Margeson, in C. J. Parise, ed., *Science and Technology of Building Seals, Sealants, Glazing, and Waterproofing*, ASTM STP 1168, ASTM, Philadelphia, Pa., 1992, pp. 22–29.
2. Y. Matsumoto, in Ref. 1, pp. 30–44.
3. D. R. Flackett, in J. M. Klosowski, ed., *Science and Technology of Building Seals, Sealants, Glazing, and Waterproofing*, Vol. 2, ASTM STP 1200, ASTM, Philadelphia, Pa., 1992, pp. 10–28.
4. G. R. Fedor, in Ref. 3, pp. 10–28.
5. J. Beech and J. L. Beasley, in Ref. 3, pp. 64–73.
6. J. M. Klosowski and D. Beers, in *Engineered Materials Handbook*, Vol. 3, ASM International, Materials Park, Ohio, 1990, p. 215.
7. D. Liles and N. Shephard, in Ref. 3, pp. 280–291.
8. R. E. Meyer, in J. I. Kroschwitz, ed., *Encyclopedia of Polymer Science and Engineering*, 2nd ed., Vol. 15, John Wiley & Sons, New York, 1989, pp. 131–145.
9. R. P. Deltieure, "Polyurethane Adhesives and Sealants," *Proceedings of Caulks and Sealants Short Course II*, The Adhesive and Sealant Council, Rosemont, Ill., 1992.
10. R. Evans and R. Greene, in J. Panek, ed., *Building Seals and Sealants*, ASTM STP 606, ASTM, Baltimore, Md., 1976, pp. 112–133.
11. J. F. Regan, "Urethane Formulations," *Proceedings of Caulks and Sealants Short Course*, The Adhesive and Sealant Council, Dallas, Tex., 1990.
12. A. R. Fiorillo, in Ref. 3, pp. 292–298.
13. M. A. Shewin, "High Performance Acrylic Latex Sealants," in Ref. 9.
14. **U.S. Pat. 4,626,567** (Dec. 2, 1986), W. T. Chang (to Beecham Home Improvement Products Inc.).
15. M. V. Newton, S. D. Halbe, and G. D. Krysiak, "Butyl Sealants: Formulating, Developing, Processing," in Ref. 11.
16. M. M. Schlechter, *Adhesives Age*, **36**(4), 30–31 (1993).
17. *Sealant Specifications Compendium*, The Adhesive and Sealant Council, Washington, D.C., 1988.

Richard A. Palmer
Jerome M. Klosowski
Dow Corning Corporation

SEASONINGS.

See FLAVORS AND SPICES

SEAWEED COLLOIDS.

See GUMS

SEDIMENTATION

Sedimentation has been defined as "the separation of a suspension into a supernatant clear fluid and a rather dense slurry containing a higher concentration of solid" (1). This definition is too broad. It does not specify the external field of acceleration, gravitational, centrifugal, magnetic, or electrostatic, that causes the separation. There is also a possible ambiguity whether the suspension is gaseous or liquid. Herein sedimentation is restricted to the most common definition, ie, the gravitational settling of solids in liquids (see also MINERAL RECOVERY AND PROCESSING; SEPARATION, CENTRIFUGAL SEPARATION; SEPARATION, MAGNETIC SEPARATION).

The uses of sedimentation in industry fall into the following categories: solid–liquid separation (2); solid–solid separation; particle-size measurement by sedimentation (3); and other operations such as mass transfer, washing, etc. In solid–liquid separation, the solids are removed from the liquid either because the solids or the liquid are valuable or because these have to be separated before disposal. If the primary purpose is to produce the solids in a highly concentrated slurry, the process is called thickening. If the purpose is to clarify the liquid, the process is called clarification. Usually, the feed concentration to a thickener is higher than that to a clarifier. Some types of equipment, if correctly designed and operated, can accomplish both clarification and thickening in one stage (see also EXTRACTION, LIQUID SOLID).

In solid–solid separation, the solids are separated into fractions according to size, density, shape, or other particle property (see SIZE REDUCTION). Sedimentation is also used for size separation, ie, classification of solids (see SEPARATION, SIZE SEPARATION). One of the simplest ways to remove the coarse or dense solids from a feed suspension is by sedimentation. Successive decantation in a batch system produces closely controlled size fractions of the product. Generally, however, particle classification by sedimentation does not give sharp separation (see SIZE MEASUREMENT OF PARTICLES).

In particle-size measurement, gravity sedimentation at low solids concentrations (<0.5% by vol) is used to determine particle-size distributions of equivalent Stokes' diameters in the range from 2 to 80 μm. Particle size is deduced from the height and time of fall using Stokes' law, whereas the corresponding fractions are measured gravimetrically, by light, or by x-rays. Some commercial instruments measure particles coarser than 80 μm by sedimentation when Stokes' law cannot be applied.

Sedimentation is also used for other purposes. For example, relative motion of particles and liquid increases the mass-transfer coefficient. This motion is particularly useful in solvent extraction in immiscible liquid–liquid systems (see EXTRACTION, LIQUID LIQUID). An important commercial use of sedimentation is in continuous countercurrent washing, where a series of continuous thickeners is used in a countercurrent mode in conjunction with reslurrying to remove mother liquor or to wash soluble substances from the solids. Most applications of sedimentation are, however, in straight solid–liquid separation.

Principles

Gravity Settling of a Single Particle. If a particle moves relative to the fluid in which it is suspended, the force opposing the motion is known as the drag force. Knowledge of the magnitude of this force is essential if the particle motion is to be studied. Conventionally, the drag force F_D is expressed according to Newton:

$$F_D = C_D A \frac{\rho v^2}{2}$$

(1)

where v is the particle–fluid relative velocity, ρ is the fluid density, A is the area of the particle projected in the direction of the motion, and C_D is a coefficient of proportionality known as the drag coefficient. Newton assumed that the drag force results from the inertia of the fluid and that C_D is then constant.

Dimensional analysis (qv) shows that C_D is generally a function of the particle Reynolds number:

$$Re_p = \frac{vx\rho}{\mu}$$

(2)

where x particle size and μ is liquid viscosity. The form of the function depends on the regime of the flow. This relationship for rigid spherical particles is shown in Figure 1. At low Reynolds numbers, under laminar flow conditions when viscous forces prevail, C_D can be determined theoretically from Navier-Stokes equations; the solution is known as Stokes' law:

$$F_D = 3 \pi \mu v x$$

(3)

This is an approximation that gives the best results for $Re_p \rightarrow 0$. The upper limit of its validity depends on the error that can be accepted. The usually quoted limit for the Stokes' region of $Re_p = 0.2$ is based on the error of ca 2%.

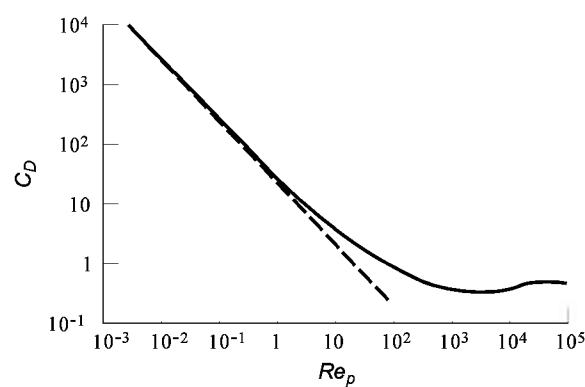


Fig. 1. Drag coefficient vs particle Reynolds number for spherical particles where () corresponds to the theoretical value of $C_D = 24 / Re_p$ (eq. 4).

Elimination of F_D between equations 1, 2, and 3 gives another form of Stokes' law, as shown in Figure 1 as a straight line.

$$C_D = \frac{24}{Re_p} \quad Re_p < 0.2 \tag{4}$$

For Reynolds numbers > 1000 , the flow is fully turbulent. Inertial forces prevail and C_D becomes constant and equal to 0.44, the Newton region. The region in between $Re_p = 0.2$ and 1000 is known as the transition region and C_D is either described in a graph or by one or more empirical equations.

In solid–fluid separation, the fine particles are most difficult to separate, ie, Re_p is low, almost inevitably below 0.2, owing to low values of α and v . Therefore, only the Stokes' region has to be considered.

A single particle settling in a gravity field is subjected primarily to drag force, F_D ; gravity force, mg ; and buoyancy, $(mg) \frac{\rho}{\rho_s}$; which have to be in equilibrium with the inertial force:

$$m \frac{dv}{dt} = (mg) - (mg) \frac{\rho}{\rho_s} - F_D \tag{5}$$

assuming positive downward forces, where m is particle mass, g is gravity acceleration, ρ_s is particle density, and t is time. Equation 5 can be solved, assuming Stokes' law:

$$v(t) = \frac{gx^2(\rho_s - \rho)}{18\mu} \left[1 - \exp\left(-\frac{t}{x^2} \frac{18\mu}{\rho_s - \rho}\right) \right] \tag{6}$$

This relationship is exponential with respect to time t and with increasing time quickly approaches equation 7, where v_g is known as the terminal settling velocity under gravity.

$$v_g = \frac{gx^2(\rho_s - \rho)}{18\mu} \tag{7}$$

The terminal velocity in the case of fine particles is approached so quickly that in practical engineering calculations the settling is taken as a constant velocity motion and the acceleration period is neglected. Equation 7 can also be applied to nonspherical particles if the particle size x is the equivalent Stokes' diameter as determined by sedimentation or elutriation methods of particle-size measurement.

Settling of Suspensions. As the concentration of the suspension increases, particles get closer together and interfere with each other. If the particles are not distributed uniformly, the overall effect is a net increase in settling velocity because the return flow caused by volume displacement predominates in particle-sparse regions. This is the well-known effect of cluster formation which is significant only in nearly monosized suspensions. For most practical widely dispersed suspensions, clusters do not survive long enough to affect the settling behavior and, as the return flow is more uniformly distributed, the settling rate steadily declines with increasing concentration. This phenomenon is referred to as hindered settling and can be theoretically approached from three premises (4): as a Stokes' law correction by introduction of a multiplying factor; by adopting effective fluid properties for the suspension different from those of the pure fluid; and by determination of bed expansion using a modified Carman-Kozeny equation. These three approaches yield essentially identical results:

$$\frac{v_p}{v_g} = \epsilon^2 f(\epsilon) \tag{8}$$

where v_p is the hindered settling velocity of a particle, v_g is the terminal settling velocity of a single particle as calculated from Stokes' law (eq. 7), ϵ is volume fraction of the fluid (voidage), and $f(\epsilon)$ is a voidage function, which for Newtonian fluids has different forms, depending on the theoretical approach adopted. The differences between the available expressions for $f(\epsilon)$ are not great and are frequently within the experimental accuracy. The most important forms are as follows. From the Carman-Kozeny equation (5),

$$f(\epsilon) = \frac{\epsilon}{10(1 - \epsilon)} \tag{9}$$

from Brinkman's theory (6), applied to Einstein's viscosity equation (7),

$$f(\epsilon) = \epsilon^{2.5} \tag{10}$$

and from the well-known Richardson and Zaki equation (8),

$$f(\epsilon) = \epsilon^{2.65} \tag{11}$$

For irregular or nonrigid particles, eg, flocs, the Einstein constant (2.5) and the Richardson and Zaki exponent (2.65) can be considerably larger than for spheres.

Strictly speaking, these correlations apply only to the cases where flocculation is absent, such as for coarse mineral suspensions. Suspensions of fine

particles, because of the very high specific surface of the particles, often flocculate and therefore show different behavior. With increasing concentration, C , of such suspensions, at a particular concentration, C_1 , an interface is observed that becomes sharper at $C > C_1$. The slurry is then said to be in the zone-settling region. The particles below the interface, if the size range is not more than 6:1, settle en masse, that is, all at the same velocity irrespective of size. There are two possible reasons for this: either the flocs become similar in size and settle at the same velocity, or they are joined and fall as a web. Interestingly enough, the settling rates of the interface, and of the solids below it, of many practical suspensions can still be described by the Richardson and Zaki equation (eq. 11) but the value of v_g has to be determined by extrapolation of the experimental log-linear plot for $\epsilon = 1$. The value of this intercept has in fact been used for indirect size measurement of the flocs. The slope of the plot determines the exponent.

The concentration C_c at which zone-settling is first observed depends on the material and its state of flocculation and no guidance can be given. Addition of flocculating agents (qv) or dispersants (qv) drastically changes this concentration and only experimental evaluation can yield its value. If the concentration is increased still higher, a point is reached when the flocs become significantly supported mechanically from underneath as well as hydraulically and the suspension is then known to be in compression or compression settling. The solids in compression continue to consolidate. The consolidation rate depends not only on the concentration but also on the structure of the solids, which in turn depends on the pressure and flow conditions. This is a complex problem closely related to cake filtration and expression (see FILTRATION). At intermediate concentrations between those of zone settling and fully established uniform compression, channeling is sometimes observed, which particularly occurs in slowly raked large-scale thickeners. Under these conditions, a coarser structure of pores becomes interconnected in the form of channels.

Most authors who have studied the consolidation process of solids in compression use the basic model of a porous medium having point contacts which yield a general equation of the mass-and-momentum balances. This must be supplemented by a model describing filtration and deformation properties. Probably the best model to date (ca 1996) uses two parameters to define characteristic behavior of suspensions (9). This model can be potentially applied to sedimentation, thickening, cake filtration, and expression.

Coagulation and Flocculation. Both coagulation and flocculation are classical pretreatment methods used to increase the effective particle size, thereby improving sedimentation settling rates. Although these two terms are often used interchangeably, coagulation is sometimes defined as agglomeration of the primary particles into particles up to 1 mm in diameter. Flocculation, on the other hand, not only agglomerates particles but also interconnects them by means of long-chain molecules of the flocculating agent into giant loose flocs up to 1 cm in size (see FLOCCULATING AGENTS). The term flocculation is used here to include coagulation as defined. Chemical agents create favorable conditions for flocculation by neutralization of surface charges and thus reduce interparticle repulsion. Mineral coagulants are in the form of electrolytes, such as alum or lime, whereas flocculation agents are mostly synthetic polyelectrolytes of high molecular weight. Development of the latter group since the 1970s has resulted in a remarkable improvement in sedimentation equipment. Such agents are relatively expensive and the correct dosage has to be carefully optimized. Overdosage is not only uneconomical but may inhibit the flocculation process and cause operating problems. Because surface charges are also affected by pH, control of this variable is essential in pretreatment.

Although reduction or elimination of the repulsion barrier is a necessary prerequisite of successful flocculation, the actual flocculation in such a destabilized suspension is effected by particle-particle collisions. Depending on the mechanism that induces the collisions, the flocculation process may be either perikinetic or orthokinetic.

Perikinetic flocculation is the first stage of flocculation, induced by the Brownian motion. It is a second-order process that quickly diminishes with time and therefore is largely completed in a few seconds. The higher the initial concentration of the solids, the faster is the flocculation.

The well-known DLVO theory of colloid stability (10) attributes the state of flocculation to the balance between the van der Waals attractive forces and the repulsive electric double-layer forces at the liquid-solid interface. The potential at the double layer, called the zeta potential, is measured indirectly by electrophoretic mobility or streaming potential. The bridging flocculation by which polymer molecules are adsorbed on more than one particle results from charge effects, van der Waals forces, or hydrogen bonding (see COLLOIDS).

The flocculation rate is determined from the Smoluchowski rate law which states that the rate is proportional to the square of the particle concentration by number; inversely proportional to the fluid viscosity, and independent of particle size.

Orthokinetic flocculation is induced by the motion of the liquid obtained, for example, by paddle stirring or any other means that produces shear within the suspension. Orthokinetic flocculation leads to exponential growth which is a function of shear rate and particle concentration. Large-scale one-pass clarifiers used in water installations employ orthokinetic flocculators before introducing the suspension into the settling tank (see WATER, MUNICIPAL WATER TREATMENT). Scale-up of orthokinetic flocculators, generally in the form of paddle devices, is based on the product of mean velocity gradient and time, for a constant volume concentration of the flocculating particles.

Another type of flocculation results from particle-particle collisions caused by differential settlement. This effect is quite pronounced in full-size plants where large rapidly falling particles capture small particles that settle more slowly.

A third type of flocculation, mechanical syneresis, has been defined (11). This is the process of shrinkage and densification of loose and bulky flocs through uneven application of local fluctuating mechanical forces leading to exudation of the liquid from the floc. It is achieved by slowly stirring the blanket zones with rotating paddles in sludge-blanket clarifiers. Pellet-like flocs can be produced by this process allowing higher overflow rates than those obtained using conventional blanket clarifiers. So far, the process has been applied successfully to only a few suspensions in water treatment.

Settling Tests

For the simplest case of particulate clarification where no flocculation takes place during settling, ie, either the flocculation process is completed before entering the settling tank or the suspension is entirely nonflocculant, the basic test is the so-called short-tube procedure (12). A sample settles in and is decanted from a large measuring cylinder in order to evaluate the settling rate, ie, the specific overflow rate that produces a satisfactory clarity of the overflow. The long-tube procedure is designed for systems where flocculation (or deflocculation) takes place during settling, thus the settling tank's performance depends not only on the specific overflow rate but also on the residence time in the tank. Tests are conducted in a vertical tube that is as long as the expected depth of the clarifier, under the ideal assumption that a vertical element of a suspension which has been clarified maintains its shape as it moves across the tank.

When the overflow clarity is independent of overflow rate and depends only on detention time, as in the case for high solids removal from a flocculating suspension, the required time is determined by simple laboratory testing of residual solid concentrations in the supernatant versus detention time under the conditions of mild shear. This determination is sometimes called the second-order test procedure because the flocculation process follows a second-order reaction rate.

The design of the sludge-blanket clarifiers used primarily in the water industry is based on the jar test and a simple measurement of the blanket expansion and settling rate (12). Different versions of the jar test exist, but essentially it consists of a bank of stirred beakers used as a series flocculator to optimize the flocculant addition that produces the maximum floc-settling rate. Visual floc-size evaluation is usually included.

The critical settling flux essential for evaluating the requirement of the zone-settling layer area of a gravity thickener is measured either using the Coe and Clevenger method (13) or the simpler Talmage and Fitch procedure (12). The former consists of a series of settling tests in a measuring cylinder where the initial settling rates of a visible interface within the settling suspension are measured for different initial solids concentrations. The Talmage and Fitch procedure is simpler because it requires only one test at any concentration, providing it is in the zone-settling regime. Theoretically, the two methods should give an identical critical settling flux and therefore identical pool areas, but this is not so in practice. Usually the Coe and Clevenger method leads to underdesign of the thickener area, whereas the Talmage and Fitch procedure leads to overdesign. For highly flocculant slurries, the area requirement of the compression layer may exceed that of the zone-settling layer; the compression zone also has a depth requirement. A laboratory test for the latter, employing a multiple batch upflow test for compression-zone evaluation, has been described (9). This test is by no means generally accepted and its reliability remains to be demonstrated.

Design Methods

The simplest case of a gravity settling tank without coagulation or flocculation in clarification applications, ie, when removing small amounts of solids, is based on the identical principle as laminar settling chambers for cleaning gases (14). The grade-efficiency curve $G(x)$ is given by equation 12:

$$G(x) = \frac{v_g A}{Q} \tag{12}$$

where v_g is the terminal settling velocity of a particle of size x (see eq. 7), Q is the feed flow rate which is equal to overflow rate, and A is the plan area of the tank. Equation 12 was derived assuming laminar flow in the tank and no end effects. The tank area is the only design parameter affecting the theoretical separational performance irrespective of the shape or depth of the pool. Equation 12 can be expressed in terms of the more conventional dimensionless groups:)

$$Stk Fr = G(x) \tag{13}$$

if Stokes' law is assumed for the particle-settling velocity and therefore the Stokes' number, Stk , is defined as follows:

$$Stk = \frac{x^2 \Delta \rho}{18 \mu} \frac{Q}{AH} \tag{14}$$

and Froude number, Fr , as in equation 15:

$$Fr = \frac{Hg A^2}{Q^2} \tag{15}$$

where H is a characteristic dimension of the pool, eg, the height, which in equation 13 cancels; $\Delta \rho$ is particle–fluid density difference; μ , liquid viscosity; and g , the acceleration of gravity.

Scale-up can be calculated with the help of equation 12, based on a simple specific overflow rate model: for the same performance, the flow rate is proportional to area, or vice-versa. Because of the long residence time involved during which the feed solids must remain constant, it is difficult to measure the grade-efficiency curve of a settling tank, particularly on a large scale. It is therefore more practical to measure the specific overflow rate (or overflow volume flux) Q/A that gives satisfactory overflow clarity from simple settling tests (12). If the clarification is not completely particulate, ie, when flocculation takes place and has not been completed before the suspension enters the settling tank, the overflow clarity depends not only on the overflow rate but also on the detention time. The scale-up under such conditions is based on the long-tube procedure. Sometimes the effect of the detention time is so strong that the overflow rate can be ignored and the scale-up is based on the detention procedure. For the whole range of coagulation clarifiers used predominantly in the water industry, the scale-up is usually based on the overflow rate determined from jar tests primarily designed to select the best flocculant and determine the settling rates that give a clear supernatant.

Probably more relevant to the chemical industry is the scale-up of thickeners. Thickeners are basically gravity settling tanks that, apart from producing a clear overflow, are designed to have a thick underflow with as little water content as possible. The feed into a thickener is generally more concentrated than the feed into a clarifier, and quite often exhibits zone-settling behavior because of the application of flocculants.

An operating thickener has basically three layers: the topmost clarification layer, the zone-settling layer, and the compression layer at the bottom. Each of the three layers requires a certain area. Ideally, the largest of these governs the design of the thickener. In most cases, it is the function of the clarification layer to prevent those particles that have escaped from the zone-settling layer or the feed from leaving with the overflow. This function is frequently less important than the thickening function and thus the thickener area is usually chosen on the basis of the zone-settling or compression layer requirement.

The conventional design and scale-up of thickeners operated with mineral or certain metallurgical slurries is based on the area requirement of the zone-settling layer and assumes that the compression zone only imposes a solids-detention (hence depth) demand and has no independent demand on area, with the exception of the empirical three-foot (1 m) rule, which if applicable, introduces an area demand (12). It is the basic principle of this method that the solids on their way downward from the feed layer to the underflow continuously increase in concentration from that of the feed to that of the underflow, usually determined by a time-retention test. The total solids flux (mass flow rate of solids per unit area), which different layers in the thickener are capable of accommodating, usually goes through a minimum between the feed zone and the underflow. This critical solids flux, G_c , determines the minimum design area of the thickener. The zone-settling layer does not form under these conditions and makes no depth demand. The thickener is then in no danger of overflowing solids. The area of the thickener, A , is calculated from the critical solids flux, G_c ; the feed flow rate, Q ; and concentration, C_f , using a simple mass balance, which assumes a complete separation of all feed solids.

$$G_c A = Q C_f \tag{16}$$

Thus the design and scale-up of the thickeners in this category centers around the determination of the critical solids flux, G_c . This value cannot be estimated from the primary properties of the particulate system because of the unpredictable effect of flocculation. It must be obtained experimentally from large-scale or pilot-scale thickeners. However, with a few exceptions, this method is impracticable because of the scale and cost of such experiments. If the settling velocity of the solids is assumed to be a function of concentration only, ie, $v = v(c)$, then this function should be unique for a given suspension and should be the same in batch-settling and continuous operations. This is the basis of the conventional Coe and Clevenger, and Talmage and Fitch methods which only differ in the way in which G_c is determined from experimental tests. The latter method requires simpler tests. According to Coe and Clevenger, if the function of $v = v(c)$ is known, the critical flux G_c corresponds to the minimum on the total flux G curve which is as follows:

$$G = v \left/ \left(\frac{1}{C} - \frac{1}{C_v} \right) \right. \tag{17}$$

where the underflow concentration, C_v , is determined by time-detention tests. The thickness of the compression zone is also determined by time-detention tests, assuming that the solids concentration reached in the compression layer in a batch test after a given time is the same as in the compression layer of a thickener.

Equipment

Sedimentation equipment can be divided into batch-operated settling tanks and continuously operated thickeners or clarifiers. The operation of the former is simple. Whereas use has diminished, these are employed when small quantities of liquids are to be treated, for example in the cleaning and reclamation of lubricating oil (see RECYCLING, OIL). Most sedimentation processes are operated in continuous units.

Clarifiers. The largest user of clarifiers is probably the water-treatment industry. The conventional one-pass clarifier uses horizontal flow in circular or rectangular vessels (Fig. 2) with feed at one end and overflow at the other. The feed is preflocculated in an orthokinetic (paddle) flocculator

which often forms an integral part of the clarifier. Settled solids are pushed to a discharge trench by paddles or blades on a chain mechanism or suspended from a traveling bridge (see WATER, INDUSTRIAL WATER TREATMENT).

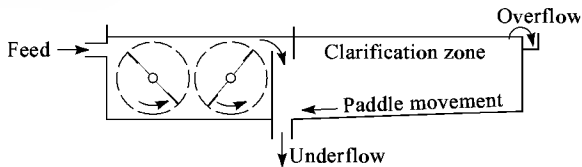


Fig. 2. Schematic diagram of a rectangular basin clarifier having an orthokinetic flocculator where the feed is mixed with a flocculant.

Circular basin clarifiers are most commonly fed through a centrally located feed well. The overflow is led into a trough around the periphery of the basin. The bottom gently slopes to the center and the settled solids are pushed down the slope by a number of motor-driven scraper blades that revolve slowly around a vertical center shaft. This design closely resembles a conventional thickener. Like thickeners, circular clarifiers can be stacked in multitray arrangements to save space. Some juice clarifiers are also arranged in this way.

Circular raking mechanisms are sometimes also used in square basins with horizontal flow across the basin. Such designs have to incorporate supplementary rake arms that reach into the corners of the square vessel (Fig. 3).

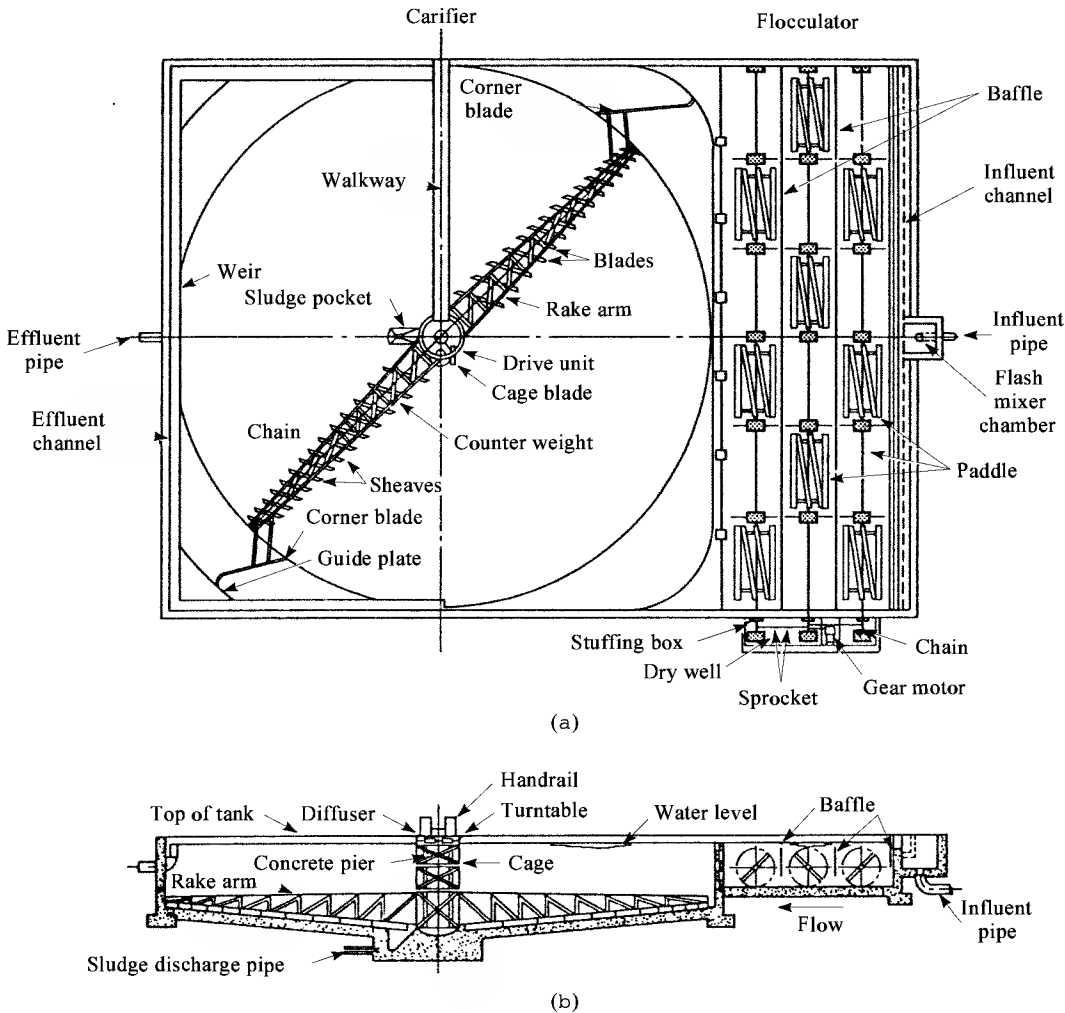


Fig. 3. The Dorco Flocculator-Squarex clarifier combination (cross-flow arrangement): (a) plan and (b) sectional elevation.

Courtesy of Dorr-Oliver Inc.

The conventional one-pass clarifier is designed for the lowest specific overflow rate (flow per unit area of liquid surface), which is usually 1–3 m³/h depending on the degree of flocculation. These clarifiers can be started and stopped without difficulty.

Newer designs incorporate some vertical flow and combine flocculation, gravity and inertial clarification, and solids recirculation. Because such units achieve higher overflow rates, they are referred to as rapid settling or high rate clarifiers. Figure 4 gives an example in this category.

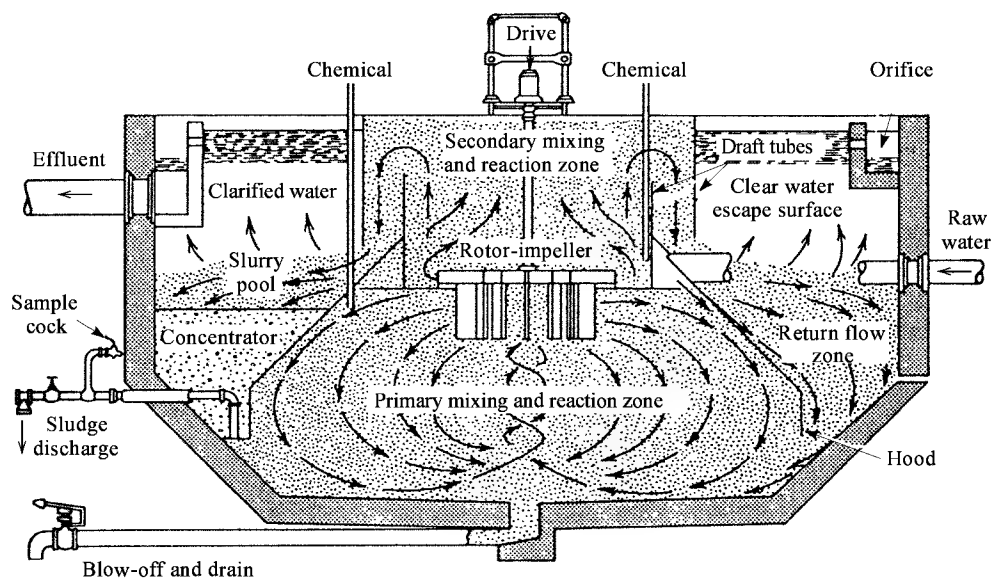


Fig. 4. Functional diagram of an accelerator where the slurry pool is indicated by shaded areas.

Courtesy of Infilco Fuller Co./General American Transportation Corp.

Flocculation is accelerated and higher overflow rates are achieved by external or internal recirculation of settled solids into the feed which leads to the collection of fine particles by interception. Addition of conditioned fine sand to the feed induces separation by differential sedimentation, and sometimes increases overflow rates to 6–8 m/h.

Design and operation of recirculation systems can be complicated. Problems are avoided by using a sludge-blanket clarifier, in which feed enters below a blanket of accumulated and flocculated solids which become fluidized in the zone-settling regime by the upflowing feed. Feed solids are trapped in the blanket. The solids content of the blanket continuously increases and part must be bled off in order to maintain the mass balance.

Sludge-blanket clarifiers are available in flat-, trough-, and hopper-bottom types. The hopper-bottom vertical-flow clarifier shown in Figure 5 achieves rise rates of 1–6 m/h in wastewater applications. It is a 3–4 m deep, 60° triangular or circular hopper tank, with feed introduced through a downward-pointing inlet at the bottom of the tank. The flocculated feed suspension is clarified by passage through the blanket and overflows into decanting troughs that are usually 1–1.5 m above the blanket, to allow for blanket-level variations with feed flow rate. The blanket can be continuously bled off through a submerged weir-type regulator and then thickened in a conical concentrator, or the clarifier can be periodically shut down to allow settlement bleeding.

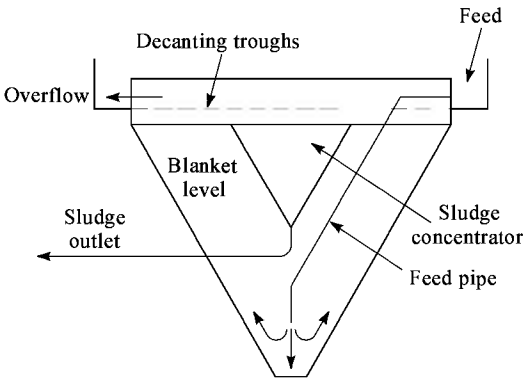


Fig. 5. Hopper-tank sludge-blanket clarifier.

Sludge-blanket clarifiers are difficult to start up because the first blanket must be established, and large-scale units require extensive excavation. Sizes range from 600 × 600 mm to 50 × 50 m. Precipitation and crystallization can be carried out in similar hopper-designed units, having overflow rates of 80 m/h or higher.

A combination of gravity settling and slow cyclonic action, first patented in 1983 (15), is available commercially. Known as the hydrodynamic or vortex-induced separator, it is used for excess storm water treatment or for the separation of grits and sands from raw sewage and other liquids. The principle of this separator (Fig. 6) is gravity settling in a circular vessel from a slow vortex flow induced by a tangential feed. The bottom of the vessel is conical and has a large included angle from which the settled solids are swept toward the central underflow discharge. The radially inward flow at the bottom, which is responsible for the sweeping action, is often referred to as the "tea cup" effect. This is caused by rotating eddies that form in cyclonic flows in vessels having wide included angles or flat bottoms.

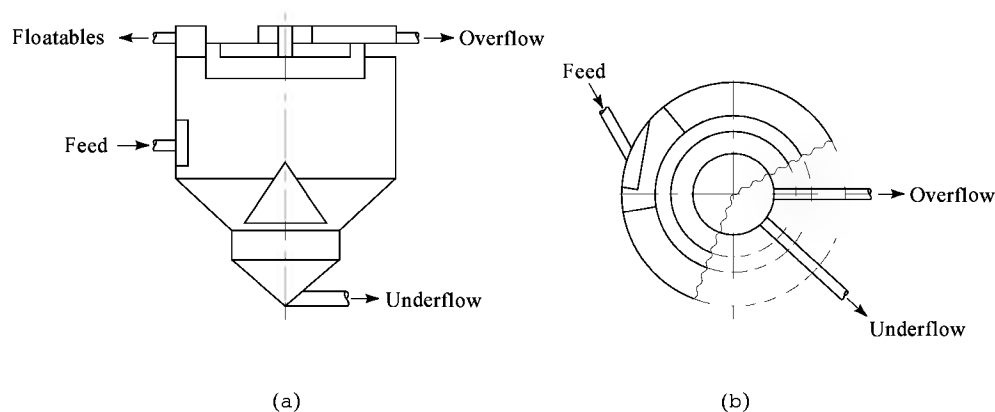


Fig. 6. Schematic diagram of the vortex-induced separator: (a) section and (b) plan. Courtesy of Hydro Research & Development Ltd.

The vessel design features a Chinese hat-like conical core stopper above the underflow sump, which is there to prevent the vortex from reaching the latter and reentraining the settled solids. The core stopper is also believed to stabilize and locate the vortex flow in the vessel. Overflow from the vessel is through a wide cylindrical insert through the lid, similar to a vortex finder in a hydrocyclone (16), and an optional provision can be made for collecting any floatables in a float trap.

Units are available in stainless steel or protected mild steel, often prefabricated, up to 12.5 m in diameter, capable of processing $>5 \text{ m}^3/\text{s}$ depending on the separation efficiency required. When the separator is used for classification of granular solids, smaller-diameter ($<4 \text{ m}$) units are used, separating nearly all particles coarser than $\sim 150 \mu\text{m}$.

Thickeners

The most common thickener is the circular basin type shown in Figure 7. After treatment with flocculant, the feed stream enters the central feed well which dissipates the stream's kinetic energy and disperses it gently into the thickener. The feed finds its height in the basin where its density matches the density of the inside suspension and spreads out at that level. Solids concentration increases downward in an operating thickener giving stability to the process. The settling solids and some liquid move downward. The amount of the latter depends on the underflow withdrawal rate. Most of the liquid moves upward and into the overflow which is collected in a trough around the periphery of the basin.

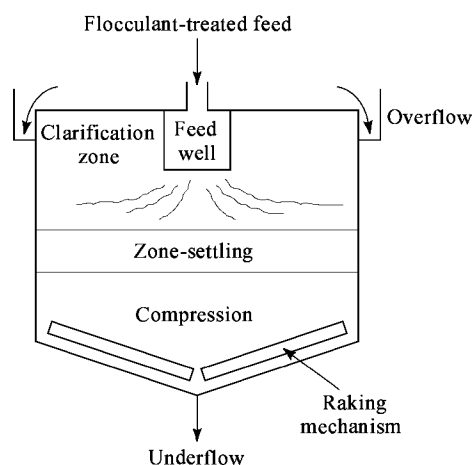


Fig. 7. The circular basin continuous thickener.

A typical thickener has three operating layers: clarification, zone-settling, and compression. Frequently, the feed is contained in the zone-settling layer which theoretically eliminates the need for the clarification zone because the particles would not escape through the interface. In practice, however, the clarification zone provides a buffer for fluctuations in the feed and the sludge levels.

The most important design dimensions of a thickener are pool area and depth. The pool area is chosen to be the largest of the three layer requirements. In most cases, only the zone-settling and compression layer requirements need to be considered. However, if the clarity of the overflow is critical, the clarification zone may need the largest area. As to the pool depth, only the compression layer has a depth requirement because the concentration of the solids in the underflow is largely determined by the time detention and sometimes by the static pressure. Thickness of the other two layers is governed only by practical considerations.

Thickeners are widely used, particularly in the mineral processing industry and in wastewater treatment. Typical applications include thickening of alumina red mud, alumina hydrate, coal tailings, copper middlings and concentrate, magnesium hydroxide, china clay (kaolin), phosphate slimes, potash slimes, pulp-mill wastes, and gas-washing effluents. In hydrometallurgical installations, thickeners are employed for the separation of dissolved components from leached residues in countercurrent washing configurations, eg, in copper, uranium, alumina, and precious metals production (see also FLOTATION).

The most widely used conventional thickeners in metallurgical and mineral processing applications give solids fluxes (mass flow rate of solids per unit area) in the range of $0.011\text{--}0.022 \text{ kg}/(\text{m}^2\text{s})$ (17). The conventional thickeners are constructed of steel ($\sim 25 \text{ m}$ dia) or concrete ($<200 \text{ m}$ dia). The floor is usually sloped toward the underflow discharge in the center. Large thickeners have earth bottoms. Raking mechanisms turning slowly around the center column consolidate the solids in the compression layer and facilitate the discharge of solids. Smaller basins can be covered for conservation of heat or to prevent freezing.

The center-drive mechanism and feed launder are usually supported by a walkway that extends across one-half or the whole diameter of the basin. Devices having drive mechanisms and rakes supported by a truss across the diameter of the thickener are referred to as bridge machines. The bridge thickeners usually do not exceed 25–45 m in diameter. In thickeners with larger diameters, the drive mechanism is supported by a central column or pier and the rates are driven and supported by a drive cage. The sediment is discharged into an annular trench around the bottom of the column.

In even larger caisson thickeners, the central column is sufficiently large to accommodate a discharge pump at the bottom. The discharge passes through the column and along the access walkway above the basin. Caisson thickeners can be built having diameters of up to 200 m. These often have an earth bottom.

The rake arms are driven by fixed connections or dragged by cables or chains suspended from a drive arm that is rigidly connected to the drive

mechanism. The rake arms are connected to the bottom of the central column by a special arm hinge that allows both horizontal and vertical movements. This arrangement lifts the rakes automatically if the torque becomes excessive. The drive arm can be attached below the suspension level or, if scaling is a problem, above the basin.

The traction thickener includes a traction mechanism where the movement of the rake is supplied by a single long arm pivoted around the center column and driven by a trolley that moves on a peripheral rail around the basin. Such units have diameters of 60–130 m.

Improved application of flocculating agents has resulted in several other high capacity thickeners capable of handling fluxes up to 0.19–0.38 kg (m²s). A good example is the deep-cone thickener developed by the National Coal Board (NCB, U.K.) (18,19). It is based on the deep-cone vessel used for the processing of coal and metallurgical ores since the turn of this century (see COAL). As can be seen in Figure 8, the vessel is equipped with a slow-turning stirring mechanism (2 rpm) which enhances flocculation in the upper part and acts as a rake in the lower section. The unit is used for densification of froth-flotation tailings at overflow rates from 6.5–10 m³ h⁻¹; the final discharge contains 25–35 wt % moisture. Other commercial deep-cone thickeners are of particular advantage, as is the NCB unit, where the final underflow density is increased by the large weight of solids above the discharge point, eg, with flocculated clays.

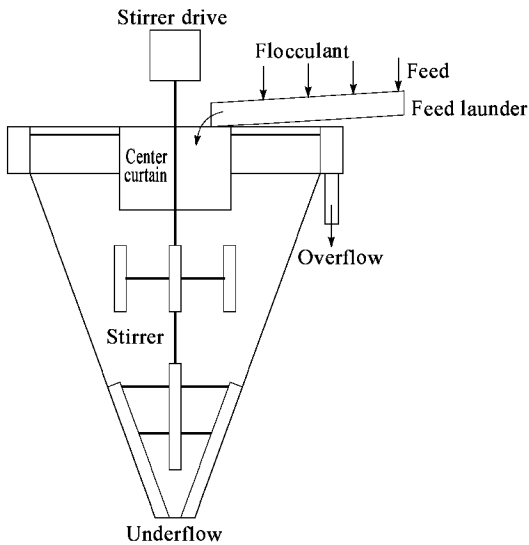


Fig. 8. The NCB deep-cone thickener.

An unraked version of the deep-cone thickener was developed for the thickening of red mud, the insoluble residue from caustic digestion of bauxite in the Bayer process, in a countercurrent decantation system (20). The deep-cone units are the last two in a nine-stage system. Whereas these are smaller in diameter for the same solids throughput, they are able to produce thicker underflows than the remaining seven conventional, raked thickeners. This greater thickening translates into fewer stages needed for the same washing and thus represents a further space saving.

There are two U.S. thickeners based on feeding the slurry under the settling–solids interface (sludge blanket) in a way similar to the sludge-blanket clarifiers described. These also offer other features designed to accelerate flocculation and thus increase the capacity. In the Eimco Hi-Capacity thickener the feed is introduced from the top through a hollow shaft which incorporates flocculant addition and a mixing device. The feed is then directed into the established sludge blanket under the sludge line which partially submerges a set of inclined settling plates. The settled solids are moved by a conventional raking mechanism at the bottom of the basin.

The principle of another high capacity unit, the Enviro-Clear thickener, is shown schematically in Figure 9. Here the flocculated feed enters vertically from the bottom and is directed horizontally at a controlled velocity into the sludge blanket by an impingement plate. A number of other arrangements are also available. The feed can, for example, enter from the top through a center well surrounding the rake driveshaft. A glass window allows visual observation of the sludge line. Available unit sizes range from 4–18 m with typical overflow rates from 2.4–14.4 m³ h⁻¹. The applications include sugar (qv) and paper (qv) production, and mineral processing (see METALLURGY, EXTRACTIVE METALLURGY).

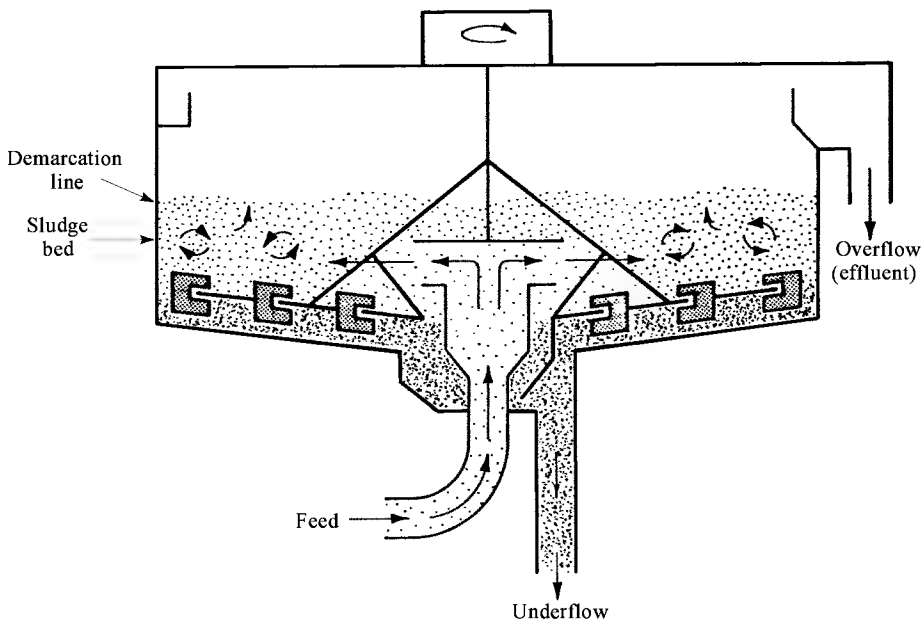


Fig. 9. The Enviro-Clear thickener.

Courtesy of Amstar Corp.

Stacking of sedimentation units in vertical arrangements increases the capacity per unit area. Multiple compartment or tray thickeners consist of two

or more conventional thickener compartments up to 35 m in diameter stacked vertically. Each compartment has a set of raking arms operating from a rotating central shaft common to the whole stack. The compartments are used either in series or parallel.

Another development in this category is the Swedish Lamella thickener (Fig. 10) which consists of a number of inclined plates stacked closely together. The flocculated feed enters the stack from the side feed box. The flow moves upward between the plates while the solids settle onto the plate surfaces and slide down into the sludge hopper underneath where they are further consolidated by vibration or raking. Even distribution of the flow through the plates is assisted by flow-distribution orifices placed in the overflow exit. In theory, the effective settling area is the sum of the horizontal projected areas of all plates. In practice, only ca 50% of the area is utilized. When treating sticky sludges, the whole Lamella pack can be vibrated intermittently or continuously to assist the sliding motion of the solids down the plates. In some instances, the plates are corrugated instead of flat or they are replaced by tube bundles in the tube settler; these tubes are square and have a cross section of ca 50 × 50 mm² and are 950 mm long. The Lamella thickener and the tube settler are used in the treatment of coal, gas-scrubber effluents, fly ash, leach solutions, and many other applications. In coal concentration, the typical overflow rates range from 1.7 to 2.9 m/h.

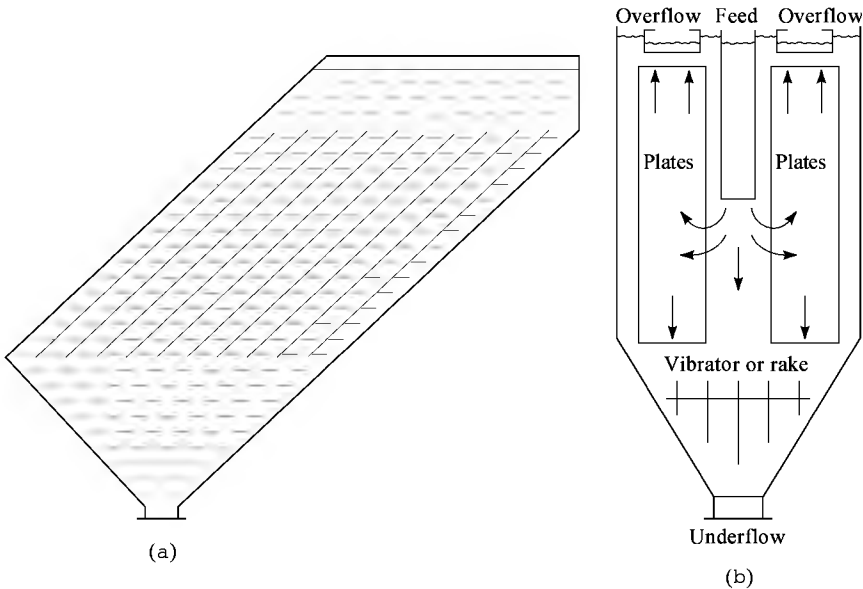


Fig. 10. The Lamella thickener: (a) side elevation and (b) end elevation.

NOMENCLATURE	
Symbol	Definition
A	area of particle projected in direction of motion; plan area of settling tank or thickener
C	solids concentration
C_D	drag coefficient
C_f	feed solids concentration
C_i	underflow concentration
$F(\epsilon)$	voidage function
F_D	drag force
Fr	Froude number
G	total flux
g	gravity acceleration
$G(x)$	grade efficiency
G_c	critical solids flux
H	height
m	particle mass
Re_p	particle Reynolds number
Stk	Stokes' number
t	time
Q	flow rate
v	particle–fluid relative velocity
$v(c)$	velocity as a function of concentration
v_g	terminal settling velocity
v_p	hindered settling velocity of a particle
$v(t)$	velocity as a function of time
x	particle size
Δ_p	particle–fluid density difference
ϵ	voidage
μ	liquid viscosity
ρ	liquid (fluid) density
ρ_s	solids (particle) density

BIBLIOGRAPHY

"Sedimentation" in *ECT* 1st ed., Vol. 12, pp. 126–145, by W. A. Lutz and D. C. Gillespie, The Dorr Co.; in *ECT* 2nd ed., Vol. 17, pp. 785–808, by F. L. Bosqui, Consultant; in *ECT* 3rd ed., Vol. 20, pp. 559–575, L. Svarovsky, University of Bradford.

1. G. G. Brown and Associates, *Unit Operations*, 7th Printing, John Wiley & Sons, Inc., New York, 1960.

2.

L. Svarovsky, ed., *Solid–Liquid Separation*, 3rd ed., Butterworths, London, 1990.

3.

T. Allen, *Particle Size Measurement*, 3rd Ed., Chapman and Hall, London, 1981.

4.

P. H. T. Uhlherr, *Hindered Settling in Non-Newtonian Fluids—an Appraisal*, CHER 75-1, Dept. of Chemical Engineering, Monash University, Melbourne, Australia, Jan. 1975.

5.

P. C. Carman, *Flow of Gases through Porous Media*, Butterworths, London, 1956.

6.

H. C. Brinkman, *Appl. Sci. Res.* **A1**, 27 (1947); **A1**, 81 (1948); **A2**, 190 (1949).

7.

A. Einstein, *Ann. Phys. Leipzig* **19**, 289 (1906); **34**, 591 (1911).

8.

J. F. Richardson and W. N. Zaki, *Chem. Eng. Sci.* **3**, 65 (1954).

9.

P. Kos, paper presented at *The Second World Filtration Congress 1979*, Olympia, London, Sept. 18–20, 1979, pp. 595–603.

10.

Ref. 2, Chapt. 4.

11.

T. Ide and K. Katoka, in Ref. 9, pp. 377–385; *Filtr. Sep.*, 52 (Mar. Apr. 1980).

12.

D. B. Purchas, ed., *Solid–Liquid Separation Equipment Scale-Up*, Upland Press Limited, London, 1977.

13.

H. S. Coe and C. H. Clevenger, *Trans. Inst. Min. Eng. (London)* **55**, 356 (1916).

14.

L. Svarovsky, *Solid–Gas Separation*, Elsevier, Amsterdam, the Netherlands, 1981.

15.

Brit. Pat. 2,082,941B (July 6, 1983), B. Smisson (to Hydro Research and Development).

16.

L. Svarovsky, *Hydrocyclones*, H. H. Rinehart & Winston, London, 1984.

17.

J. M. Keane, *World Min.*, 44 (Nov. 1979).

18.

J. Abbott and co-workers, paper presented at *The International Coal Preparation Congress*, Paris, France, 1972, paper 20E VI.

19.

J. Abbott, *Filtr. Sep.* **16**(4), 376 (July–Aug. 1979).

20.

J. L. Chandler, *IChemE Symposium Solids/Liquids Separation Practice*, Leeds, U.K., Apr. 2–5, 1984, pp. 177–182.

General References

R. D. Paradis, *Filtr. Sep. Technol.* **7**, 175–179 (Apr. 1993).

M. Marunczyn and T. J. Laros, *Filtr. Sep. Technol.* **7**, 180–188 (Apr. 1993).

O. E. Albertson, *Filtr. Sep. Technol.* **7**, 198–203 (Apr. 1993).

"The Lamella Gravity Settler; Theory, Design and Experience," *Filtr. Sep. Technol.* **7**, 204–211 (Apr. 1993).

G. Butt, R. Holditch, and A. S. Ward, *Proceedings of the Filtech Conference, Karlsruhe, Germany, Oct. 19–21, 1993*, Vol. 2, Filtration Society, Horsham, U.K., 1993, pp. 33–42.

L. Svarovsky, in L. R. Weatherley, ed., *Engineering Processes for Bioseparations*, Butterworth-Heinemann, Oxford, U.K., 1994, Chapt. 5, pp. 110–134.

Ladislav Svarovsky

Consultant, Engineers and Fine Particle Software

SELENIUM AND SELENIUM COMPOUNDS

Selenium

Selenium [7782-49-2], Se, atomic no. 34, atomic wt 78.96, lies between sulfur and tellurium in Group 16 (VIA) and between arsenic and bromine in Period 4 of the Periodic Table (1–4). Its outer electronic configuration is $3d^{10} 4s^2 4p^4$, ie, the three inner electron shells of selenium are completely filled. Strikingly similar to sulfur in most of its chemistry, important selenium oxidation states are -2 , 0 , $+2$, $+4$, and $+6$. The $+2$ state is not known to occur in nature. Selenium exists in various allotropic modifications and forms many inorganic and organic compounds, often analogous and isomorphous with their sulfur equivalents. Selenium was discovered in 1817 and its name derived from *selene*, the Greek goddess of the moon.

Physical Properties. The six stable isotopes of selenium are ^{74}Se [13981-33-4], ^{76}Se [13981-32-3], ^{77}Se [14681-72-2], ^{78}Se [14833-16-0], ^{80}Se [14681-54-0], and ^{82}Se [14687-58-2]. These occur naturally in the approximate abundances of 0.87, 9.02, 7.58, 23.52, 49.82, and 9.19%, respectively. A number of artificial radioactive isotopes have been prepared by neutron activation. One of these, the γ -emitting ^{75}Se [14265-71-5], which has a half-life of 120.4 d, is a diagnostic tool in medicine (see MEDICAL IMAGING TECHNOLOGY; RADIOACTIVE TRACERS).

Solid selenium has several allotropic forms, including an amorphous one resembling plastic sulfur. The stable form at ordinary temperatures, ie, the gray or hexagonal selenium, is the most dense and is semimetallic in appearance. Selenium crystallizes in a hexagonal lattice with $a = 0.4366$ nm and $c = 0.4954$ nm. The electrical conductivity, which makes gray selenium useful in photoelectrical and photochemical applications, is low in the dark but increases several hundredfold on exposure to light. Crystalline red selenium exists in two monoclinic forms obtained by evaporation of carbon disulfide extracts of amorphous red selenium. The α -monoclinic form, where $a = 0.9054$ nm, $b = 0.9083$ nm, $c = 1.1060$ nm, and $\beta = 90.81^\circ$, has a unit cell formed of four puckered Se_8 ring molecules. The β -monoclinic form, where $a = 1.285$ nm, $b = 0.807$ nm, $c = 0.931$ nm, and $\beta = 93.13^\circ$, is also made up of puckered Se_8 [13494-81-0] rings. Amorphous selenium exists in black and red forms. Black amorphous selenium is vitreous and is formed by rapid cooling of liquid selenium; red amorphous selenium is colloidal and is formed in reduction reactions. Heating and/or catalysts transform all the amorphous forms and both monoclinic crystalline forms to gray selenium. Liquid selenium is black in the bulk and brownish red in thin films. The liquid probably contains chains and rings of variable numbers of atoms. Selenium vapor is also complex in nature. The most abundant species are Se_2 [12185-17-0], Se_3 [12597-28-3], Se_4 [12597-30-7, 20721-13-5], Se_7 [12597-32-9], and Se_8 . Se_2 is an important species at 900°C, whereas at 2000°C the vapor is mainly monatomic. The saturated vapor pressure is given (5), when T is in K, by the following:

$$\log P_{\text{kPa}} = 7.2355 - 5010.7/T$$

To convert P_{kPa} to $\log P_{\text{mm Hg}}$, subtract 0.8751 from the constant. The saturated vapor pressure increases from 133 Pa (1 mm Hg) at 343.7°C to 101.3 kPa (760 mm) at 685.4°C, the boiling point of selenium.

Some physical constants for selenium are given in Table 1. More extensive data and many sources are available (1–5). For a selenium atom, the covalent radius is ca 0.115 nm, the electron affinity for two electrons is ca -2.33 eV, ie, energy absorbed, and the first ionization potential is 9.75 eV.

Table 1. Physical Constants of Selenium^a

Property	Value
melting point, °C	217
boiling point, °C	ca 685
heat of fusion, trigonal liquid, kJ mol ^b	5.2–5.4
heat of vaporization, kJ mol ^b	59.7
heat of combustion, at 298 K, kJ mol ^b	225.1
heat capacity, J (g K) ^b	
trigonal	24.52
vitreous	25.627
liquid	29.288
thermal conductivity, W (m K)	248.1
thermal expansion coefficient ^c , °C ⁻¹	3.24×10^{-5} – 7.5×10^{-5}
viscosity, mPa s (cP)	
at 220°C	221
360°C	70
density, g cm ³	
trigonal at 298 K	4.819
monoclinic	4.4
liquid at 490 K	4.05
vitreous	4.285
standard reduction potential, V	
$\text{Se} + 2 e^- \rightarrow \text{Se}^{2-}$	0.78
$\text{Se} + 2 \text{H}^+ + 2 e^- \rightarrow \text{H}_2\text{Se}(\text{aq})$	0.36
surface tension, liquid, mN m (dyn cm)	
at 220°C	105.5
310°C	95.2
electronegativity, Pauling scale	2.4

^a Refs. 2 and 4.
^b To convert J to cal, divide by 4.184.
^c Value depends on form.

Chemical Properties. The chemical properties of selenium are intermediate between those of sulfur and tellurium. Whereas selenium reacts with active metals and gains electrons to form ionic compounds containing the selenide ion, Se²⁻, selenium forms covalent compounds with most other substances. The oxidation states in elemental and combined forms are as follows (5): -2 in disodium selenide [1313-85-5], Na₂Se; -1 in disodium diselenide [39775-49-0], Na₂Se₂; 0 in Se₈; +1 in diselenium dichloride [10025-68-0, 21317-32-8], Se₂Cl₂; +2 in selenium dichloride [14457-70-6], SeCl₂; +4 in sodium selenite [10102-18-8], Na₂SeO₃; and +6 in sodium selenate [13410-01-0], Na₂SeO₄.

Selenium combines with metals and many nonmetals directly or in aqueous solution. The selenides resemble sulfides in appearance, composition, and properties. Selenium forms halides by reacting vigorously with fluorine and chlorine, and less so with halogen compounds bromine and iodine. It does not react with pure hydrogen fluoride or hydrogen chloride but decomposes hydrogen iodide to liberate iodine and form hydrogen selenide [7783-07-5], H₂Se. Selenium combines with oxygen yielding a number of oxides, the most stable being selenium dioxide [7446-08-4], SeO₂. Under proper conditions, selenium forms selenides with hydrogen, carbon, nitrogen, phosphorus, and sulfur. Crystalline selenium does not react with water, even at 150°C.

Selenium remains unaffected by dilute sulfuric acid or hydrochloric acid, but dissolves in a nitric–hydrochloric acid mixture, concentrated nitric acid (qv), and concentrated sulfuric acid. It is oxidized by ozone and solutions of alkali metal dichromates, permanganates, and chlorates and calcium hypochlorite. Selenium dissolves in strong alkaline solutions yielding selenides and selenites. It forms selenocyanates, MSeCN, with alkali metal cyanides, MCN, as well as many inorganic and organic derivatives of the corresponding acid, HSeCN [13103-11-2]. Selenium also dissolves in alkali metal sulfites, M₂SO₃, forming selenosulfates, M₂SeSO₃, and, because tellurium does not undergo this reaction, this method can be used to separate the two elements. Selenium mixes in all proportions with sulfur and tellurium forming a continuous series of solid solutions and alloys.

In many reactions, selenium is an oxidant as well as a reductant. Strong oxidants convert selenium dioxide and its derivatives to the hexavalent state. Although hexavalent selenium compounds are oxidants, these are less active and difficult to reduce. Selenium salts resemble the corresponding sulfur and tellurium salts in behavior.

Selenium also forms a large number of organic compounds. Of special interest are the oxidizing and reducing actions of selenium and its compounds. Chemical reactions are described in detail elsewhere (1–9). Organic reactions are reviewed in References 10 and 11. The organic chemistry and biochemistry of selenium is also available (12–16).

Occurrence. The occurrence of selenium is discussed in References 17–19. At 0.68 atom per 10,000 atoms Si, selenium is the 30th most abundant element. Selenium is widely dispersed in igneous rocks probably as selenide minerals; in volcanic deposits where it substitutes for some of the sulfur; in hydrothermal deposits, where it is associated isothermally with silver, gold, antimony, and mercury; and in massive sulfide and porphyry copper deposits, where it appears in large quantities but only in small concentrations. In sedimentary rocks such as sandstones, carbonaceous siltstones, phosphorite rocks, and limestones, selenium is syngenetic, ie, it was introduced during deposition, probably by adsorption on precipitated ferric hydroxide. Like vanadium, phosphorus, arsenic, and antimony, selenium has been reported in sedimentary iron ores in amounts larger than its average abundance in the crust.

The estimated selenium content of the oceans is only ca 0.5 ppb, and only a small fraction of the selenium is transported into the sea by weathering and erosion (20). The principal species in the seas is SeO²⁻₄. Adsorption by some marine organisms contributes to the removal of selenium from seawater. Thus selenium often, though not always, occurs in the pyrite and marcasite of sedimentary formations as well as in soils derived from them. Selenium is particularly concentrated in soils of the drier regions, eg, the North American great plains from Mexico to the prairie provinces of Canada and westward to the Pacific Ocean at California, but especially in Wyoming and South Dakota, and in localities in Colombia, Ireland, Israel, and the People's Republic of China. It is also found in much smaller quantities in Argentina, Venezuela, Bulgaria, Algeria, Morocco, Australia, and some regions of the former USSR, as shown by analyses of locally grown crops. Selenium occurs in coals in 0.5–12 ppm concentrations, probably associated with pyrite and marcasite (21). Crude oil from different locations contains variable amounts of selenium, usually less than 0.5 ppm.

There are areas (22) where selenium levels in the soil are very low; these include regions of volcanic activity like that adjacent to the Cascade mountains in the Pacific Northwest states of the United States and the central north island of New Zealand. There, because the heat of eruption volatilized the selenium, the residual soil parent material is virtually devoid of selenium. Other areas of low soil-selenium reflect leaching of selenium out of the top soil, as in the Canterbury plain on New Zealand's south island. Areas of selenium deficiency have negative implications for animal and human health.

Biosphere. Selenium occurs in alkaline soils chiefly as selenates which, when water soluble, are readily available to plants. In acid soils, selenides and to some degree elemental selenium are prevalent but little is available to plants. The concentration of selenium in most soils is between 0.01 and 2.0 mg/kg. However, levels as high as 1200 mg/kg have been reported in seleniferous areas. High selenium soils have developed from Cretaceous shales such as those found in the Dakotas and Wyoming in the United States, whereas low selenium soils include those derived from sedimentary rocks (northeastern U.S.) and volcanic ash (Pacific Northwest U.S.). The topsoils can be enriched by certain native plants that accumulate high selenium concentrations from the seleniferous soils and earth formations. These plants, which require selenium for their growth, include about 24 species and varieties of *Astragalus* (milk vetch), *Machaeranthera* (woody aster), *Haplopappus* (goldenweed), and *Stanleya* (Princes' plume). Other species, eg, *Atriplex* (salt bush) and *Grindelia* (gumweed), absorb moderately large quantities when growing in soils with high available selenium concentrations. These so-called secondary selenium absorbers and the primary selenium indicators aid in locating seleniferous regions. Some selenium-tolerant plants may contain up to 1.5 wt % selenium.

Indicator plants generally have an offensive odor, which varies with the selenium concentration. Other vegetable matter grown on seleniferous soils may have a sufficiently high selenium content to be toxic when ingested by animals or humans. Apart from appearance in these seleniferous plants, selenium has been considered as a variable contaminant. Selenium is a necessary micronutrient in living organisms, needed by humans as well as animals (see MINERAL NUTRIENTS).

Data on selenium in water are limited. The drinking water content is usually less than 1 µg/L, and seldom exceeds the 50 µg/L upper limit established in 1993 by the U.S. EPA (23). It may be higher in wells in seleniferous areas, and markedly higher in some river waters where irrigation drainage from seleniferous soil contains up to 2680 µg/L.

Deposits. Selenium forms natural compounds with 16 other elements. It is a main constituent of 39 mineral species and a minor component of 37 others, chiefly sulfides. The minerals are finely disseminated and do not form a selenium ore. Because there are no deposits that can be worked for selenium recovery alone, there are no mine reserves. Nevertheless, the 1995 world reserves, chiefly in nonferrous metals sulfide deposits, are ca 70,000 metric tons and total resources are ca 130,000 t (24). The principal resources of the world are in the base metal sulfide deposits that are mined primarily for copper, zinc, nickel, and silver, and to a lesser extent, lead and mercury, where selenium recovery is secondary.

Manufacture and Recovery. Electrolytic copper refinery slimes are the principal source of selenium and its sister element, tellurium, atomic numbers 34 and 52, respectively. Electrolytic copper refinery slimes are those constituents in the copper anode which are not solubilized during the refining process and ultimately accumulate in the bottom of the electrolytic tank. These slimes are periodically recovered and processed for their metal values. Slimes generated by the refining of primary copper, copper produced from ores and concentrates, generally contain from 5–25% selenium and 2–10% tellurium.

Selenium occurs in the slimes as intermetallic compounds such as copper silver selenide [12040-91-4], CuAgSe; disilver selenide [1302-09-6], Ag₂Se; and Cu_{2-x}Se_x [20405-64-5], where x < 1. The primary purpose of slimes treatment is the recovery of the precious metals gold, silver, platinum, palladium, and rhodium. The recovery of selenium is a secondary concern. Because of the complexity and variability of slimes composition throughout the world, a number of processes have been developed to recover both the precious metals and selenium. More recently, the emphasis has switched to the development of processes which result in early recovery of the higher value precious metals. Selenium and tellurium are released in the later stages. Processes in use at the primary copper refineries are described in detail elsewhere (25–44).

Soda Ash Roasting. Some of the first processes to recover selenium on a commercial basis were based on roasting of copper slimes with soda ash to convert both selenium and tellurium to the +6 oxidation state. Figure 1 shows flow sheets for two such processes. Slimes are intensively mixed with sodium carbonate, a binder such as bentonite, and water to form a stiff paste. The paste is extruded or pelletized and allowed to dry. Care in the preparation of the extrudates or pellets is required to ensure that they have sufficient porosity to allow adequate access to the air required for oxidation.

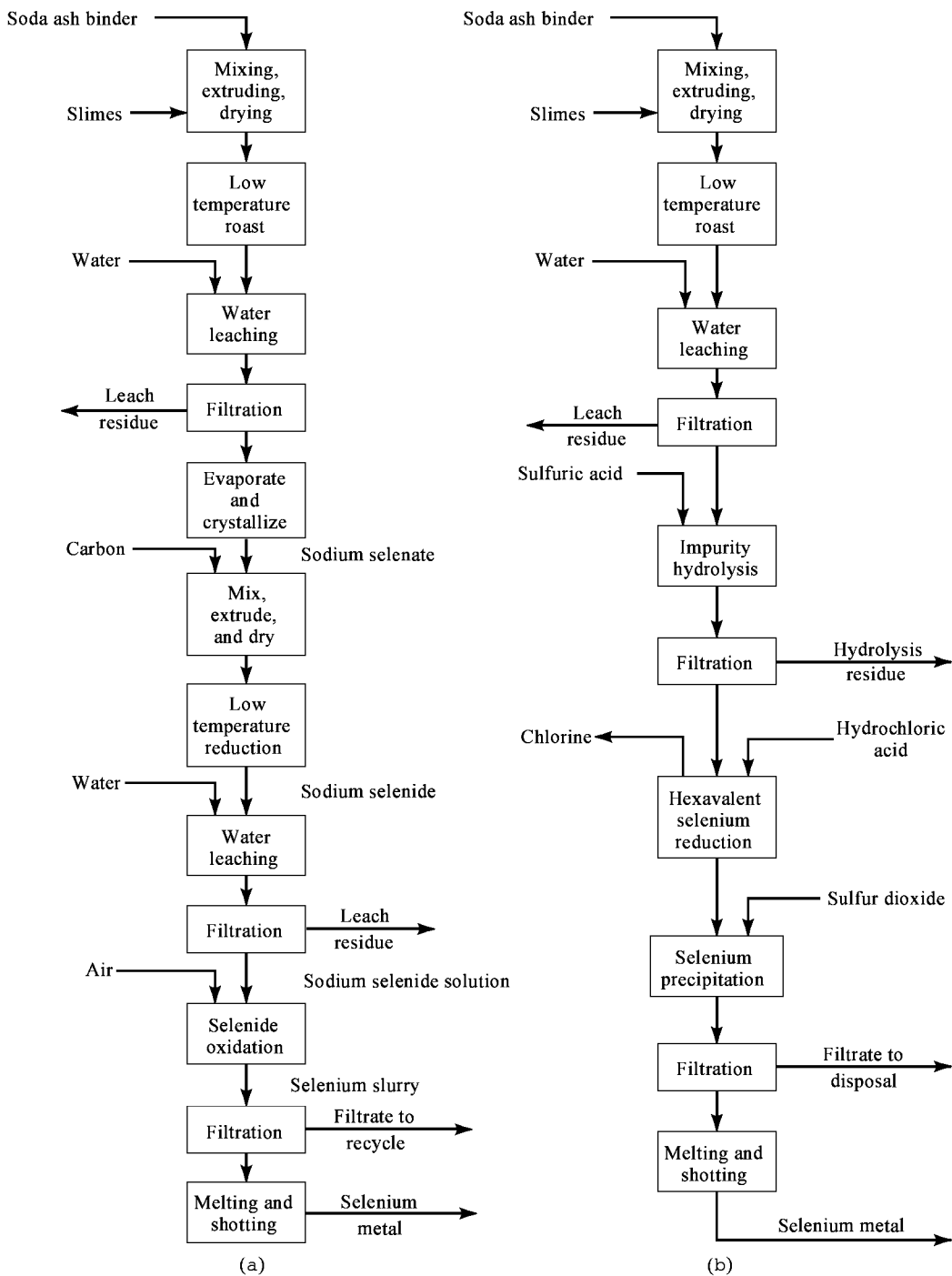


Fig. 1. Recovery of selenium by soda ash roasting of slimes: (a) process 1 and (b) process 2.

Roasting occurs between temperatures of 530–650°C. Virtually no volatilization of selenium or tellurium takes place during roasting. Conversion of both elements to the hexavalent form is complete.

The roasted pellets or extrudes are ground and leached in water. The hexavalent selenium dissolves as sodium selenate [13410-01-0], Na₂SeO₄. Sodium tellurate, being highly insoluble in the now very strongly alkaline solution, remains in the residue. The separation between selenium and tellurium is readily achieved, provided all tellurium is oxidized to the hexavalent state.

Selenium Recovery. There are two processes commonly employed for reducing selenium from solution. There are many variations on these processes. In the first process, employed commercially for many years, selenium is leached from the slimes in the form of the hexavalent sodium selenate. It is recovered from solution by crystallization (qv) and the crystalline sodium selenate is mixed with charcoal. Under controlled conditions of heating, the Na₂SeO₄ is reduced to sodium selenide. The sodium selenide cake is leached with water to form a typically liver-red solution of sodium selenide which is readily oxidized to the elemental form by blowing air through the solution. This precipitation technique allows the recycling of much of the solution, a significant advantage considering the particularly severe restrictions placed on the discharge of selenium-bearing solutions.

In the second process, reduction of the hexavalent selenium is accomplished using concentrated hydrochloric acid or ferrous iron salts catalyzed by chloride ions as the reductant. This process was developed during the 1940s in Germany and is still apparently employed at a variety of refineries for dealing with hexavalent selenium. It generates a large effluent stream of iron chlorides containing small quantities of selenium which can be burdensome to discharge or control. Additionally, the solution is extremely corrosive and considerable care must go into the selection of process equipment to avoid severe corrosion problems.

Alkaline Autoclaving. A variety of research experiments have been performed on the oxidation of slimes under pressure in an alkaline solution (32,35,40). Although many different process conditions are stipulated, a temperature of approximately 200°C is generally preferred when using sodium hydroxide concentrations on the order of 100–500 g/L, depending on the concentration of selenium in the slimes and solids of the slimes slurry employed. Oxygen partial pressures vary from 172–1720 kPa (25–250 psig). Reaction time is generally greater than four hours, and can be as much as 12 hours to achieve acceptable selenium extraction. The reactions for selenium and tellurium are shown in the following equations.



Conversion of tellurium to the hexavalent form is complete. The extent of selenium oxidation to the hexavalent form is a function of temperature,

alkalinity, and oxygen pressure. Total conversion of tellurium to the hexavalent form ensures its virtually complete insolubility in the alkaline leach liquor, thereby allowing an essentially quantitative separation from the soluble selenium compounds. The advantages of the alkaline pressure leach process include its comparatively low corrosivity, no volatile selenium loss, no scrubbing or gas cleaning circuit, and a substantially quantitative separation of tellurium and selenium.

The process is not without its disadvantages. The drawbacks include high oxygen consumption, because oxygen is used not only for the oxidation of the selenium and tellurium but for a variety of other constituents in slimes, including organic constituents introduced through the additives employed as growth modifiers in the copper refinery. Sodium hydroxide consumption may also be high, because lead sulfate in the slimes is converted to lead plumbate and silica present in the slimes from mold washes is converted to sodium silicate. In fact, substantially all metal sulfates present in the slimes are converted to sodium sulfate and their respective metal oxide, hydroxide, or sodium salts. In addition, alkaline autoclaving has the same drawbacks as soda ash roasting with regard to reduction of hexavalent selenium and tellurium compounds.

Sulfation Roasting. Acid roasting technology (Fig. 2) relies on differences in the volatility of the tetravalent oxides of selenium and tellurium at roasting temperatures of 500–600°C to selectively volatilize selenium from slimes. Acid roasting uses sulfuric acid as the oxidant for the conversion of selenium selenides and tellurium tellurides to their respective tetravalent oxides. Typical oxidation reactions are as follow:

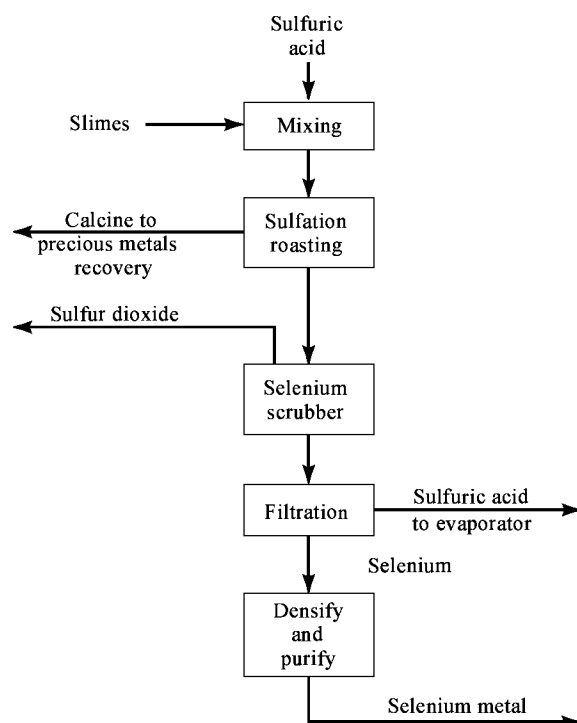
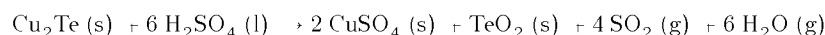
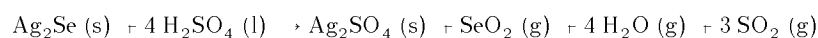
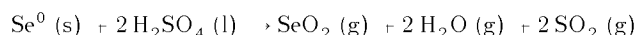
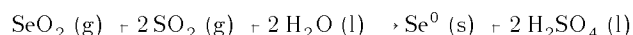


Fig. 2. Recovery of selenium by sulfation roasting.

It is apparent from these equations that significant quantities of sulfur dioxide are generated. For selenium, the reaction shown for oxidation of elemental selenium reverses itself at the lower temperatures employed for water scrubbing, thus regenerating sulfuric acid. The tellurium dioxide remains in the sulfated slimes.



Selenium Recovery. Because the oxidation reactions take place well below the temperature at which significant volatilization of elemental selenium occurs, only the readily scrubbable selenium dioxide is present in the vapor phase. Thus, the scrubbing of the off-gases is highly efficient, resulting in virtually complete recovery of the selenium from the gas stream. Although some sulfuric acid is recovered in dilute form during the scrubbing, this process consumes more sulfuric acid than is recovered, because only sulfur dioxide associated with the oxidation of selenium from the elemental state to the tetravalent state is reoxidized to sulfuric acid. All sulfuric acid associated with the oxidation of tellurium, copper, and silver is irretrievably lost as sulfur dioxide.

It is generally unacceptable to emit sulfur dioxide, thus the scrubber effluent must be treated for sulfur dioxide removal. If the plant already possesses facilities for the production of sulfuric acid, this rather concentrated sulfur dioxide stream can be easily fed into the wet gas cleaning circuit and disposed of in the sulfuric acid plant. The quantity is so small that it does not put any additional burden on the sulfuric acid plant. Because no tellurium is carried over with the selenium dioxide during roasting, it is possible to produce a selenium product which can be purified to commercial grade (99.5–99.7% o).

The process is not without its difficulties. These include the rather long reaction times required for the oxidation of the selenium selenides and extensive foaming owing to sulfur dioxide liberation during the reaction. This mandates large allowances for freeboard, making the equipment large and cumbersome. Also, the equipment must be designed so that selenium formed in the scrubber does not plug scrubber nozzles or pipelines (qv) during its phase transformation from monoclinic to hexagonal crystal structure, which ordinarily occurs at about 60°C. This problem is sometimes dealt with by the introduction of lignone or similar compounds into the scrubber solution. These interdict the phase transformation but may cause serious problems downstream in producing a selenium of saleable density and purity.

Several refineries use this process which was invented by Outokumpu Oy (Finland) in the 1970s. The original batch process hearth roasted a slurry of decopperized slimes and concentrated sulfuric acid. The off-gases, containing selenium dioxide and sulfur dioxide, were scrubbed using water and a venturi jet ejector. Phelps Dodge Refining Corporation (El Paso, Texas) has converted this to a continuous process. Outokumpu has modified their process so that trays of dry decopperized slimes are contacted with a gaseous mixture of sulfur dioxide and water in an oven. Sulfur trioxide is generated *in situ* and converts the selenium to selenium dioxide which is volatilized from the trays into the overall gas stream. The gases pass through a venturi scrubber and selenium is precipitated as before.

Another variation is practiced by CCR Division of Noranda (Canada) (37) in which the selenium is vaporized from the sulfated decopperized slimes

in a top-blown rotary converter (TBRC). The advantage of using a TBRC is that further refining can be carried out in the furnace to directly produce a doré bullion containing the precious metals. In the Outokumpu process the deselenized slimes must be transferred from the trays to a separate refining furnace if bullion is to be produced.

Chlorination Processes. Both aqueous (wet) and high temperature (dry) chlorination processes have been developed for processing decopperized and detellurized slimes (29,30). Dry chlorination processes have not achieved the acceptance of wet processes because of problems in controlling temperature during fixed-bed chlorination, difficulty in separating the volatile chlorides of selenium and tellurium from those of gold and platinum metals, and loss of bed porosity owing to fusion of silver chloride. In particular, problems associated with heat transfer to avoid bed fusion, owing to the low melting point of silver chloride, are a principal barrier to the commercialization of this process.

On the other hand, wet chlorination of refinery slimes has proven to be a rapid and simple method of obtaining high extractions of selenium from slimes. A simple wet chlorination flow sheet is shown in Figure 3. Slimes chlorination per se is not a simple deselenization operation, but rather a process wherein virtually all the constituents of slimes which form soluble chlorides report as a complex solution of mixed chlorides. Thus the use of wet chlorination requires a complete change in the process to recover the metal values in slimes. The first plant to use wet chlorination of slimes was started by Kennecott (Salt Lake City, Utah) in 1995.

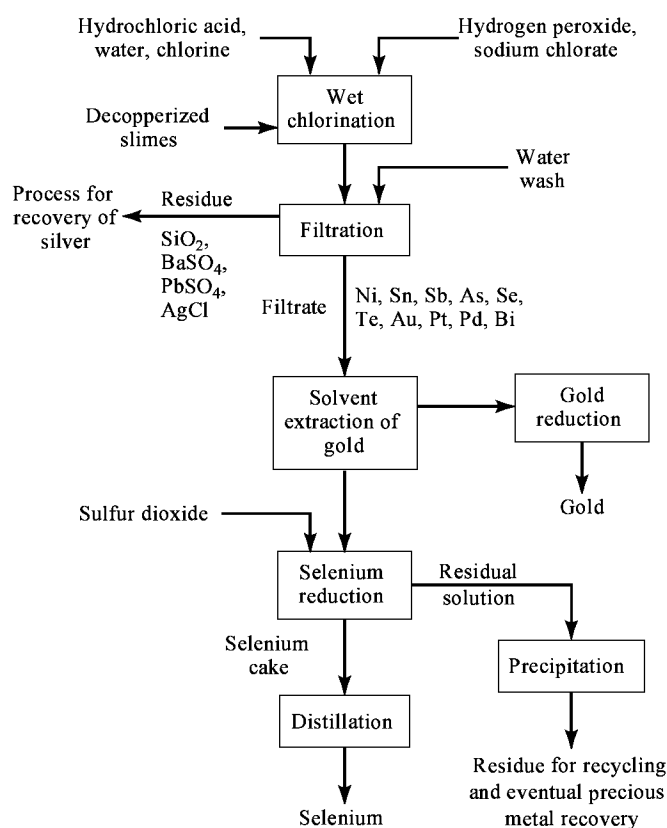
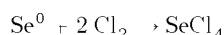
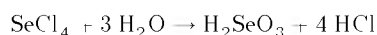


Fig. 3. Recovery of selenium by wet chlorination.

Wet chlorination is performed by sparging slimes slurried either in water or hydrochloric acid using chlorine gas, or other oxidants such as sodium chlorate or hydrogen peroxide which liberate chlorine from hydrochloric acid, at about 100°C. Under these conditions, selenium and selenides rapidly oxidize and dissolve.



Selenium tetrachloride hydrolyzes to form hydrochloric acid, the degree of hydrolysis depending on the acidity of the solution.



Chlorine may initially convert the selenium in solution to the hexavalent state, but as the hydrochloric acidity increases, reduction to the tetravalent state occurs spontaneously.

Losses of selenium and tellurium from the solution are negligible, provided the reactor is equipped with a reflux condenser. The wet chlorination is easily controlled. The reaction is rapid, allowing fast turnover of the precious metals in the slimes and yielding all the selenium and tellurium in soluble form.

Selenium and precious metals can be removed selectively from the chlorination liquor by reduction with sulfur dioxide. However, conditions of acidity, temperature, and a rate of reduction must be carefully controlled to avoid the formation of selenium monochloride, which reacts with elemental selenium already generated to form a tar-like substance. This tar gradually hardens to form an intractable mass which must be chipped from the reactor. Under proper conditions of precipitation, a selenium-precious metals product substantially free of other impurities can be obtained. Selenium can be recovered in a pure state by vacuum distillation, leaving behind a precious metals residue.

Miscellaneous. Where a copper refinery is adjacent to a lead (qv) plant it is feasible to recover the selenium in slimes by smelting them in conjunction with lead-bearing materials. Utilizing the lower temperatures needed to melt lead, the selenium is volatilized from a lead bath or cupel blown with air. The selenium is recovered from flue dust and fume by scrubbing. This is the process used by Union Minière at its Hoboken plant in Belgium.

Purifying Selenium and Tellurium. Selenium recovery processes generally yield a metal product which contains some tellurium, and, correspondingly, recovered tellurium generally contains some small amount of selenium (37,41).

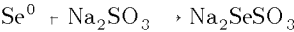
Classically, the purification of elemental selenium from tellurium is achieved by distillation because of the much greater volatility of selenium compared to tellurium. The advantages of distillation, typically at 1.0–0.255 kPa (0.015–0.037 psi) and 300–400°C, are its simplicity and the absence of emission problems, if performed under vacuum on properly dried, acid-free material. Single-stage distillation can also be carried out at atmospheric pressure at 680°C from cast-iron or iron alloy retorts. In this manner most of the occluded impurities such as sulfur dioxide, water, organic matter, halogens, sulfuric acid, and mercury are expelled to a scrubber, leaving the bulk of the nonvolatile impurities such as tellurium and other metals in the residue. Distillation from a quartz retort upgrades the quality of the selenium.

Complete separation of selenium from tellurium cannot be achieved owing to classical Raoult's law considerations and what appears to be the

formation of a Se–Te azeotrope at about 1–2 mol % selenium. A simpler and highly effective purification can be achieved by fluxing molten selenium at about 300°C in the presence of a eutectic mixture of potassium and sodium nitrates (40). At 300°C, selenium does not react significantly with the eutectic nitrate mixture; however, tellurium is oxidized, as are many other impurity metals, and passes into the molten flux layer.

This process can be readily substituted for distillation and proves equal and in some cases superior to distillation. It does not, however, remove precious metals and requires the disposal of the spent sodium–potassium nitrate slag as well as the need to treat any NO_x emissions.

In a cyclic method selenium is dissolved in hot sodium sulfite solution to form sodium selenosulfate [25468-09-1].



Tellurium and many other impurities remain undissolved. The solution is filtered and cooled to reverse the reaction and to deposit solid selenium. Deselenized liquor is recycled to the dissolution step.

Selenium purification by zone refining is not feasible. At practical zone-refining speeds, crystallization does not occur and impurities do not segregate. However, a controlled differential thermal treatment of selenium in a long vertical glass tube has been described (45). The treatment time is several weeks to several months.

Selenium Compounds

Selenium is incorporated into a large variety of commercially used compounds. The most common are listed in Table 2.

Table 2. Commercial Selenium Products

Product	CAS Registry Number	Formula	Synthetic method	Se con-tent, ^a %	Refs. ^b
barium selenite	[13718-59-7]	BaSeO ₃		29.9	
barium selenate	[7787-41-9]	BaSeO ₄		28.2	
cadmium sulfoselenide	[11112-63-3]	CdSSe		35.3	46
calcium selenite	[13780-18-2]	CaSeO ₃		47.3	
ferroselenium ^c (iron–selenium)	[1310-32-3]	FeSe	melting a mixture of powdered iron and selenium	58.6	
nickel–selenium ^c	[1314-05-2]	NiSe	melting a mixture of powdered nickel and selenium	57.2	
potassium selenocyanate	[3425-46-5]	K(SeCN)		54.9	
selenic acid	[7783-08-6]	H ₂ SeO ₄	oxidizing selenous acid with hydrogen peroxide; electrolyzing selenous acid ^d	54.5	47
selenous acid	[7783-00-8]	H ₂ SeO ₃		61.2	
selenium diethyldithio-carbamate (Selenac)	[5456-28-7]	S NC S Se	reaction of diethylamine, carbon disulfide, and selenium dioxide in alcoholic solution		48
selenium dioxide ^e	[7446-08-4]	SeO ₂	oxidizing selenium with nitric acid and evapo-rating; oxidizing selenium with air or oxygen	71.1	
selenium disulfide	[7488-56-4]	SeS ₂		55.2	
selenium oxychloride	[7791-23-3]	SeOCl ₂	chlorinating mixture of selenium and selenium dioxide dry or suspended in CCl ₄	47.6	49
selenium tetrachloride	[10026-03-6]	SeCl ₄		35.7	
selenourea	[630-10-4]	(NH ₂) ₂ CSe		64.2	
sodium selenate	[13410-01-0]	Na ₂ SeO ₄	neutralizing selenic acid with soda ash; electrolyzing sodium selenite solution ^f	41.8	50
sodium selenide	[1313-85-5]	Na ₂ Se		63.2	
sodium selenite	[10102-18-8]	Na ₂ SeO ₃	neutralizing selenous acid with soda ash	45.7	

[(C₂H₅)₂NC S]₄Se reaction of diethylamine, carbon disulfide, and selenium dioxide in alcoholic solution | | 48 || selenium dioxide^e | [7446-08-4] | SeO₂ | oxidizing selenium with nitric acid and evapo-rating; oxidizing selenium with air or oxygen | 71.1 | |
selenium disulfide	[7488-56-4]	SeS₂		55.2	
selenium oxychloride	[7791-23-3]	SeOCl₂	chlorinating mixture of selenium and selenium dioxide dry or suspended in CCl₄	47.6	49
selenium tetrachloride	[10026-03-6]	SeCl₄		35.7	
selenourea	[630-10-4]	(NH₂)₂CSe		64.2	
sodium selenate	[13410-01-0]	Na₂SeO₄	neutralizing selenic acid with soda ash; electrolyzing sodium selenite solution^f	41.8	50
sodium selenide	[1313-85-5]	Na₂Se		63.2	
sodium selenite	[10102-18-8]	Na₂SeO₃	neutralizing selenous acid with soda ash	45.7	

^a Theoretical value.

^b The patents cited are basic; improvements have been made, but little has been published.

^c Grade is crushed lump.

^d The electrolyzed product contains no selenous acid.

^e Grade is commercial and high purity.

^f The electrolyzed product contains no selenite.

Also reported as commercial products are selenium amino acid chelate, selenium ascorbate, selenium aspartate, and selenium citrate titrations, as well as selenium dibromide, selenium dichloride, selenium lysinate, and selenium sulfide bentonite.

Inorganic Compounds. Inorganic selenium compounds are similar to those of sulfur and tellurium. The most important inorganic compounds are the selenides, halides, oxides, and oxyacids. Selenium oxidation states are –2, 0, +1, +2, +4, and +6. Detailed descriptions of the compounds, techniques, and methods of preparation, and references to original work are available (1–3,5,6–10, 51–54). Some important physical properties of inorganic selenium compounds are listed in Table 3.

Table 3. Physical Properties of Inorganic Selenium Compounds^a

Compound	CAS Registry Number	Formula	Mp, °C	Bp, °C	Δ <i>H</i> _f , 25°C, kJ mol ^b
hydrogen selenide	[7783-07-5]	H ₂ Se	65.7	41.3	85.7
carbon diselenide	[506-80-9]	CSe ₂	40–45	125	155.2
selenium monochloride	[10025-68-0]	Se ₂ Cl ₂	85	127 dec	43.7
selenium monobromide	[7789-52-8]	Se ₂ Br ₂		227 dec	

selenium dichloride	[14457-70-6]	SeCl ₂			
selenium tetrafluoride	[13465-66-2]	SeF ₄	13.2	101	
selenium tetrachloride	[10026-03-6]	SeCl ₄	305	196 sub	188.3
selenium tetrabromide	[7789-65-3]	SeBr ₄	75 dec		
selenium hexafluoride	[7783-79-1]	SeF ₆	34.6	46.6 sub	1029.3
selenium dioxide	[7446-08-4]	SeO ₂	340	315 sub	– 238.5
selenium trioxide	[13768-86-0]	SeO ₃	118	180 dec	184
selenous acid	[7783-00-8]	H ₂ SeO ₃			531.3
selenic acid	[7783-08-6]	H ₂ SeO ₄	60		– 538.0
selenium oxyfluorides	[7783-43-9]	SeOF ₂	15	125–126	
	[14984-81-7]	SeO ₂ F ₂	99.5	8.4	
selenium hypofluorite	[27218-12-8]	SeOF	54	29	
selenium oxychloride	[7791-23-3]	SeOCl ₂	10.8	177.6	
selenium oxybromide	[7789-51-7]	SeOBr ₂	41.7	217	

^a Refs. 1, 2, and 4.
^b To convert J to cal, divide by 4.184.

Selenides. Selenium forms compounds with most elements. Binary compounds of selenium with 58 metals and 8 nonmetals, and alloys with three other elements have been described (55). Most of the selenides can be prepared by a direct reaction. This reaction varies from very vigorous with alkali metals to sluggish and requiring high temperature with hydrogen.

Hydrogen Selenide. The only important hydride of selenium is hydrogen selenide, H₂Se, although there is some evidence for H₂Se₂. Deuterium selenide [13536-95-3], D₂Se, has been prepared and its properties studied. Hydrogen selenide may be prepared by the action of acids or water on some metal selenides, usually aluminum selenide or iron selenide, or by passing hydrogen and selenium vapor over pumice at 177°C resulting in a ca 58% yield. Thermodynamically, hydrogen selenide is unstable at room temperature, but the rate of decomposition is very slow. The gas is colorless, flammable, highly toxic, and has an offensive odor. It irritates the mucous membranes, nose, throat, and eyes. Dry oxygen has no effect on hydrogen selenide. Moist air decomposes it, liberating elemental selenium. It burns in air with excess oxygen to SeO₂ and with insufficient oxygen, to Se. Hydrogen selenide is a strong reductant. The aqueous solution is weakly acidic, although it is stronger than aqueous acetic acid.

Binary Selenides. Most binary selenides are formed by heating selenium in the presence of the element, reduction of selenites or selenates with carbon or hydrogen, and double decomposition of heavy-metal salts in aqueous solution or suspension with a soluble selenide salt, eg, Na₂Se or (NH₄)₂Se [66455-76-3]. Atmospheric oxygen oxidizes the selenides more rapidly than the corresponding sulfides and more slowly than the tellurides. Selenides of the alkali, alkaline-earth metals, and lanthanum elements are water soluble and readily hydrolyzed. Heavy-metal selenides are insoluble in water. Polyselenides form when selenium reacts with alkali metals dissolved in liquid ammonia. Metal (M) hydrogen selenides of the M HSe type are known. Some heavy-metal selenides show important and useful electric, photoelectric, photo-optical, and semiconductor properties. Ferroselenium and nickel selenide are made by sintering a mixture of selenium and metal powder.

The carbon selenides CSe₂ [506-80-9], COSe [1603-84-5], and CSSe [5941-19-9] are not very stable, especially when exposed to light. Nitrogen selenide [12033-88-4], Se₄N₄, is an orange-red amorphous powder and is unstable. It detonates easily when scratched or when heated to 200°C. The phosphorous selenides, P₄Se₃ [1314-86-9] and the readily flammable P₂Se [12137-67-6], are known. Sulfur and selenium form homogeneous solutions in the liquid state, and these form a series of solid solutions on crystallization. Orange SeS [7446-34-6] has been prepared by precipitation. Selenium and tellurium form a continuous series of solid solutions or alloys, which probably contain different simple and mixed chains of various lengths.

Halides and Oxyhalides.

Halides. Selenium combines directly with fluorine, chlorine, bromine, and iodine, and forms the monohalides, Se₂X₂; the dihalides, SeX₂; the tetrahalides, SeX₄; and the hexafluoride, SeF₆ [7783-79-1]. The compounds are covalent and volatile. The stability decreases from the fluorides to the iodides. Although SeF₆ is stable on an electric-arc discharge, SeI₂ [81256-76-0] and SeI₄ [13465-68-4] exist in solution only. Selenium halides and sulfur halides have similar properties. The tetrahalides are more stable than the dihalides. When heated, the dihalides decompose into selenium and a tetrahalide. The chlorides and bromides are easily hydrolyzed and the hexafluoride hydrolyzes very slowly. The hydrolysis products are selenous or selenic acids and halogen acid.

Selenium tetrafluoride, SeF₄, is a colorless liquid which fumes in air. It is a powerful oxidant. It attacks silicon, phosphorus, arsenic, antimony, and bismuth; dissolves sulfur, selenium, bromine, and iodine; reacts violently with water; and attacks glass if wet. Selenium tetrachloride, SeCl₄, is a white or pale yellow crystalline substance. It forms addition compounds with metal and nonmetal chlorides, ammonia, thiourea, and amines. It is a strong chlorinating substance; it is readily hydrolyzed to selenous acid and is reduced to the monochloride and lower chlorides by sulfur and selenium. Selenium monochloride, Se₂Cl₂, is a brownish red oily liquid. It is decomposed by a number of substances and reacts with many metals and nonmetals. Selenium tetrabromide, SeBr₄, is orange-red, crystalline, hygroscopic, and readily hydrolyzed. In carbon disulfide solution and with ammonia, it forms the highly explosive Se₄N₄. Selenium monobromide is a dark red (almost black), pungent smelling, oily liquid. It is hygroscopic and is readily hydrolyzed by water and alkalies and forms addition compounds with amines and heterocyclic bases.

Oxyhalides. Selenium oxyhalides, SeOX₂, are prepared by the addition of a halogen to a dry mixture or a carbon tetrachloride suspension of selenium and selenium dioxide. Selenium oxyfluoride is conveniently prepared by treating selenium oxychloride with silver fluoride. The product is a colorless liquid with a pungent odor. It reacts with water, silica, silicon, and very violently with red phosphorus, and it forms addition compounds with hydrofluoric acid. Selenium oxychloride is also a colorless liquid with a pungent odor, is readily hydrolyzed, and fumes in moist air. It is a strong chlorinating agent and oxidant; it reacts with many metals, nonmetals, and organic substances. It dissolves many metal chlorides and chalcogenides, and it forms solvates and addition compounds. When mixed with sulfur trioxide, it dissolves alumina, chromic oxide, and rare metal oxides. Selenium oxychloride has been called the universal solvent. It has a high dielectric constant and has been used extensively as an ionizing solvent. Selenium oxybromide is an orange solid and is very easily hydrolyzed, decomposing in moist air at 50°C. It is a brominating agent with solvent properties similar to those of the oxychloride.

Oxides, Acids, and Salts.

Oxides. Selenium monoxide [12640-89-0] has only been detected spectroscopically as the gas. Selenium dioxide, SeO₂, is prepared by burning selenium in a current of air or oxygen and, optionally, by passing it over a catalyst or by oxidation with nitric acid to selenous acid followed by evaporation to dryness by heating. The compound is white and crystalline with a tetragonal structure and is yellowish green as the vapor. It dissolves easily in water forming selenous acid and less readily in acetone, alcohols, glacial acetic acid, and dioxane. Selenium dioxide is less stable than sulfur dioxide or tellurium dioxide. Like sulfur dioxide it is an acid anhydride, but unlike SO₂ it is a strong oxidant and is reduced to elemental selenium by sulfur dioxide, hydrogen, hydrogen sulfide, ammonia, phosphorus, carbon, and even organic dust particles in moist air. The reaction with sulfur dioxide does not occur without the presence of water vapor. Selenium dioxide strongly catalyzes the oxidation of nitrogen compounds in Kjeldahl digestion. The use of selenium dioxide as an oxidant in organic chemistry has been reviewed (56). In the gaseous state it is reduced by ammonia and hydrocarbons with emission of light. This gaseous reaction with ammonia was used at one time and in the production of high purity selenium. Selenium dioxide is oxidized by fluorine to SeO₂F₂ and by hydrogen

peroxide and potassium permanganate to form selenates.

Selenium trioxide, SeO₃, is white, crystalline, and hygroscopic. It can be prepared by the action of sulfur trioxide on potassium selenate or of phosphorous pentoxide on selenic acid. It forms selenic acid when dissolved in water. The pure trioxide is soluble in a number of organic solvents. A solution in liquid sulfur dioxide is a selenonating agent. It is stable in very dry atmospheres at room temperature and on heating it decomposes first to selenium pentoxide [12293-89-9] and then to selenium dioxide.

Acids and Salts. The important oxyacids are selenous acid, H₂SeO₃, and selenic acid, H₂SeO₄. Selenous acid is formed by wet oxidation of the element or by dissolving selenium dioxide in water. It is colorless, crystalline, and highly soluble in water, forming a weakly acidic solution and two series of salts: selenites and hydrogen selenites. Selenous acid is an oxidant. It is easily reduced by nascent hydrogen, hydrogen sulfide, and sulfur dioxide. It is also readily oxidized to selenic acid by fluorine, bromine, chlorine, chloric acid, ozone, hydrogen peroxide, or permanganate; the oxidation by chlorine or bromine is reversible. Electrolysis of selenous acid yields selenic acid free of selenous acid. The selenate and hydrogen selenate salts are similar in their properties to the sulfates and hydrogen sulfates. Selenic acid and selenates are stronger oxidants than sulfuric acid or sulfates.

Some of the other selenium oxyacids are permonoselenic acid [81256-77-1], H₂SeO₅; perdiselenic acid [81256-78-2], H₂Se₂O₈; and pyroselenic acid [14998-61-9], H₂Se₂O₇. Selenosulfuric acid, H₂SeSO₃, is not known but its alkali metal salts have been prepared.

Other inorganic selenium compounds include sodium selenocyanate [4768-87-0], NaSeCN, which is prepared by melting together selenium and sodium cyanide; selenocyanogen [27151-67-3] (SeCN)₂; sodium selenosulfate [25468-09-1], Na₂SeSO₃, which is prepared by dissolving selenium in aqueous sodium sulfite (acidification decomposes this compound); and selenate alums, eg, Al₂(SeO₄)₃·K₂SeO₄ [13530-59-1].

Organic Compounds. The chemical properties of organosulfur, organoselenium, and organotellurium compounds are markedly similar. Because bond stability decreases with increasing atomic number of the element, thermal stability and stability on exposure to light of all wavelengths decrease and oxidation susceptibility increases. Thus, selenium compounds often turn red when exposed to light or air. These factors as well as the unpleasant odor of many compounds and difficulties in analytical determination in the early 1900s hindered the development of organoselenium chemistry. Since the 1960s interest has grown appreciably. A large and steadily increasing number of selenium analogues of organosulfur compounds have been prepared and their properties and possible uses have been studied. They range from the simple COSe, CSSe, and CSe₂ to complex heterocyclic compounds, selenium-containing coordination compounds, and selenium-containing polymers.

Selenocysteine was identified in 1976 (57) in a protein produced by *Clostridium stricklandii*, and it is thought to be the form in which selenium is incorporated, stoichiometrically, into proteins. Studies with rats show that over 80% of the dietary selenium given them is incorporated into proteins, thus selenocysteine takes on metabolic importance. Selenoproteins having known enzymatic activities contain selenocysteine at the active sites. Two other forms of metabolic selenium are recognized: methylated selenium compounds are synthesized for excretion, and selenium is incorporated into some transfer ribonucleic acids (tRNAs) in cultured cells (58). Some of the more important seleno-compounds are listed in Table 4. Examples of simple ring compounds are shown in Figure 4.

Table 4. Types of Organoselenium Compounds

Type	Formula ^a
selenols (selenomercaptans)	RSeH
selenides	RSeR
diselenides	RSeSeR
triselenides	RSeSeSeR
selenenic acids	RSeOH
selenamides	RC(Se)NH ₂
selenocyanates	RSeCN
seleninic acids	RSeO(OH)
selenonic acids	RSeO ₂ (OH)
halides	R ₃ SeX
dihalides	R ₂ SeX ₂
trihalides	RSeX ₃
selenoxides	R ₂ SeO
selenones	R ₂ SeO ₂
selenonium compounds	R ₃ Se ⁺ X
selenoaldehydes	RCHSe
selenoketones, cyclic selenoketones	RC(Se)R
sulfenoselenides	RSSeR
selenourea [630-10-4]	(NH ₂) ₂ CSe
selenocarboxylic acids, esters, and N-derivatives of selenourea	(RNH) ₂ CSe
selenocycloalkanes	CH ₂ (CH ₂) _x Se
selenophenols	C ₆ H ₅ SeH
selenosemicarbazides	H ₂ NC(Se)NHNHR
selenium-containing N-heterocycles	
selenium-containing carbohydrates (selenogluconides, selenosugars)	
selenium-containing dyes (cyanine and noncyanine)	

^a R organic aliphatic or aromatic group.

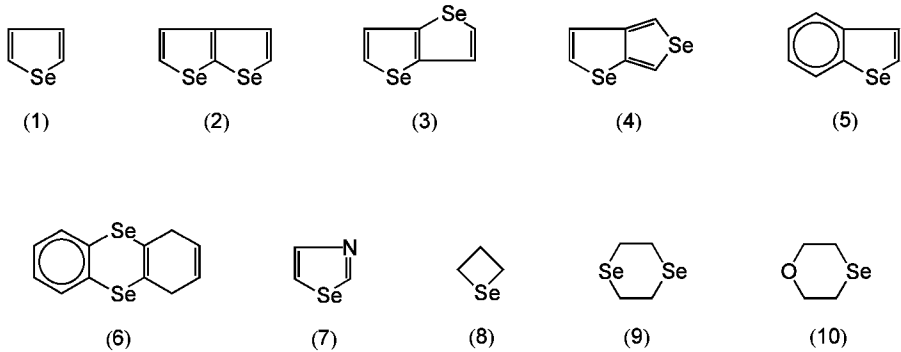


Fig. 4. Simple organic ring compounds containing selenium: selenophene [288-05-1] (1), selenolo[2,3-*b*]selenophene [250-85-1] (2), seleno[3,2-*b*]selenophene [251-49-0] (3), seleno[3,4-*b*]selenophene [250-71-5] (4), benzo[*b*]selenophene [272-30-0] (5), selenanthrene [262-30-6] (6), selenazole [288-52-8] (7), selenetane [287-28-5] (8), and cyclic ethers 1,4-diselenane [1538-41-6] (9) and 1,4-selenoxane [5368-46-7] (10).

The organoselenium compounds, their chemistry, and methods of preparation are discussed in the literature (3,4,7,11,12,14,59). Selenium-containing polymers are of interest because of their semiconducting and plastic properties. The biologically important compounds include selenoaminocarboxylic acids and derivatives, selenium-containing peptides, and selenium derivatives of pyrimidines, purines, cholines, steroids, coenzyme A, and other compounds (4,14). The biochemical and medical aspects and uses of organoselenium compounds have also been discussed (13,14,60–62).

Organoselenium compounds, such as phosphine selenides, are being evaluated in solvent extraction systems for silver and gold (63). Also, potential pharmaceuticals containing selenium have been prepared (64).

Economic Aspects

Commercial production of selenium started in the United States in 1910 when ca 4500 kg was produced. Prior to that selenium was imported from Germany. Demand remained small for several decades, but in the 1940s selenium use in electronics began to increase and selenium became, in many cases, an essential industrial material. All primary selenium producers are electrolytic copper refiners. The U.S. producers of refined selenium are Asarco Incorporated and Kennecott Copper Corporation. Two other U.S. copper refiners, Phelps Dodge Refining Corporation and Magma Copper Company, send selenium or selenium-bearing slimes outside the United States for upgrading or final refining. In Canada, the CCR Division of Noranda is a principal producer of refined selenium. Selenium producers are listed in Table 5.

Table 5. 1995 Refined Selenium Production

Country	Production, t	Selenium producer
Australia	250 ^a	Hydromet Operations Ltd., Marrickville
Belgium		Union Minière BU Hoboken, Hoboken
Canada		Inco Limited, Toronto
Chile	300	Noranda Minerals Inc., CCR Div., Montreal East, Québec
		Codelco Chile
		Corminco SA Chile, Santiago
China, P.R.	45	Enami—Empresa Nacional de Minería, Santiago
		Guangzhou Smelter, Guangzhou
		Shanghai Smelter, Shanghai
		Shenyang Smelter, Shenyang
		Zhuzhou Smelter, Zhuzhou
Finland	31	Guixi Smelter, Jiangxi
Germany	120	Outokumpu Harjavalta Metals Oy, Pori
Japan	595	Norddeutsche Affinerie AG, Hamburg
Peru	595	Japan Energy Corp., Tokyo
		Mitsubishi Materials Corp.; Tokyo Mitsui Mining & Smelting Co., Ltd.; Tokyo Nippon Rare Metal, Inc.; Yokohama Shinko Chemical Co., Ltd.; Hyogo
		Sumitomo Metal Mining Co., Ltd., Tokyo
Philippines	40	Centromin-Peru, La Oroya
Russia	50	Pacific Rare Metal Industries Inc., Quezon City, Manila
Sweden		Copper Refinery Uralektromed, Yekaterinburg
United Kingdom		Boliden Ore & Metals AB, Stockholm
United States	360	Mining & Chemical Products Ltd., Alperdon, London
Zambia	360	Asarco Inc., New York
		Kennecott Corp., Salt Lake City, Utah
		Phelps Dodge Refining Corp., El Paso, Tex.
		Zambia Consolidated Copper Mines Ltd., Kalulushi

^a Value is estimated.

Japan, Canada, and the United States accounted for 70% of the 1995 estimated world production of 2000 t (Table 5). At least 100 t of selenium was also available to Western markets from the former Soviet bloc. Selenium production is expected to rise in South America, particularly Chile, as the copper industry continues rapid expansion. A considerable amount of unrefined selenium is also shipped to Chile and the Philippines for conversion to final commercial product by either hydrometallurgical or distillation processes.

Selenium recovery from electronic, xerographic, or other scrap for sale on the open market remains small. This was less than 100 t yr in 1995. The xerographic industry recovers essentially all of the selenium it uses in photoreceptor drums by extremely efficient recycling programs (see ELECTROPHOTOGRAPHY; RECYCLING).

Like all minor metals, the unique physical and chemical properties of selenium determine its technological uses. Consumption of selenium by industry in 1995 was as follows: glass, 35%; electrical, 30%; pigments, 10%; metallurgy, 10%; agricultural biological, 5%; and miscellaneous, 10%. Two of the main markets for selenium are in decline as of the mid-1990s, resulting in a corresponding downward pressure in prices. Selenium through the 1960s to 1980s was the preferred material in xerography, initially as the element and later as arsenic triselenide. However, the use of organic photoreceptors rapidly increased in the early 1990s and as a result the amount of selenium used in xerography has declined. As of 1996, more selenium is used in laser jet printing than in photocopying and prices range from \$50–\$60 kg for high purity (99.999%) material. In addition, the use of selenium in cadmium sulfoselenide pigments is declining because of increasing environmental regulations (see PIGMENTS, INORGANIC). Declining markets have resulted in prices for commercial 99.7% grade selenium falling from \$14 kg in 1985 to \$8 kg in 1995.

Specifications and Standards

There are no official specifications for selenium, apart from those of the U.S. Government for purchases under the national stockpile plan. Some producers publish standards, and some users specify a screen analysis and maximum content of certain impurities. Four grades of selenium are available. There is the commercial or refined grade, which contains >99.5 wt% selenium and various impurities, usually <0.2 wt% tellurium, <0.1 wt% iron, <0.005 wt% lead, and <0.005 wt% copper. This grade is sold mainly as 75-μm (200-mesh) powder as well as in smaller quantities of coarser size and as lump. A second, more select pigment grade has a purity of 99.7%. The third grade, high purity selenium, contains 99.999 wt% selenium, although this figure of purity is obtained by difference and not by a direct determination. This grade is sold usually in shotted form but also in small amounts as a powder. Some users have their own specifications which reflect a particular application of selenium. Impurities, eg, mercury, tellurium, iron, arsenic, and nonferrous metals that are harmful in electronic and electrostatic uses must be <1–2 ppm each. A somewhat higher concentration of inert contaminants, eg, sodium, magnesium,

calcium, aluminum, and silicon, can be tolerated. The concentration of halogens, sulfur, and oxygen must be low, depending on the use. The fourth grade, called ultrahigh purity selenium, is claimed to contain 99.999–99.9999 wt % selenium.

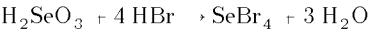
Analytical Methods

Comprehensive accounts of the various gravimetric, polarographic, spectrophotometric, and neutron activation analytical methods have been published (1,2,5,17,19,65–67). Sampling and analysis of biological materials and organic compounds is treated in References 60 and 68. Many analytical methods depend on the conversion of selenium in the sample to selenous acid, H₂SeO₃, and reduction to elemental selenium when a gravimetric determination is desired.

Several common acid treatments for sample decomposition include the use of concentrated nitric acid, aqua regia, nitric–sulfuric acids, and nitric–perchloric acids. Addition of potassium chlorate to nitric acid assists in dissolving any carbonaceous matter. Perchloric acid is an effective oxidant, but its use is hazardous and requires great care. Loss of selenium may occur on preparing the sample or during determination. For example, the practice of dissolving a sample of some metals in a nonoxidizing acid could lead to some loss of selenium as the volatile hydrogen selenide.

Organic selenium compounds and siliceous materials (rock, ore, concentrates) are fused with mixtures of sodium carbonate and various oxidants, eg, sodium peroxide, potassium nitrate, or potassium persulfate. For volatile compounds, this fusion is performed in a bomb or a closed system microwave digestion vessel. An oxidizing fusion usually converts selenium into Se(VI) rather than Se(IV).

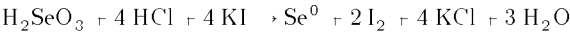
When the sample contains <0.1% selenium or if interfering substances are present, selenium may be preconcentrated by distillation from a bromine–hydrobromic acid mixture:



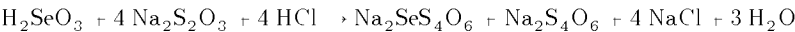
The SeBr₄ which forms is distilled from the solution leaving the interfering elements behind. The only other metallic elements that can also distill over by this procedure are arsenic, antimony, tellurium (partly), and germanium.

A number of substances, such as the most commonly used sulfur dioxide, can reduce selenous acid solution to an elemental selenium precipitate. This precipitation separates the selenium from most elements and serves as a basis for gravimetry. In a solution containing both selenous and tellurous acids, the selenium may be quantitatively separated from the latter by performing the reduction in a solution which is 8 to 9.5 N with respect to hydrochloric acid. When selenic acid may also be present, the addition of hydroxylamine hydrochloride is recommended along with the sulfur dioxide. A simple method for the separation and determination of selenium(IV) and molybdenum(VI) in mixtures, based on selective precipitation with potassium thiocarbonate, has been developed (69).

There are also numerous titrimetric methods for determining macro amounts of selenium including the iodometric, thiosulfate, and permanganate methods. In one of the several iodometric methods, the dissolved selenium is reduced from Se(IV), if an excess of potassium iodide is added:



The liberated iodine is titrated with standard sodium thiosulfate solution. In the thiosulfate method, selenous acid is treated with an excess of standard sodium thiosulfate solution:



The excess Na₂S₂O₃ is back-titrated with standard iodine solution. The permanganate method is based on the oxidation of Se(IV) to Se(VI).

From a toxicological and physiological point of view, the determination of micro amounts of selenium is becoming increasingly important. Interest in environmental and human health has promoted development of analytical techniques and methods at the trace and ultratrace levels (see TRACE AND RESIDUE ANALYSIS). The classical, highly selective and sensitive 3,3′-diaminobenzidine method is used widely in the spectrophotometric determination of low levels of selenium. Some instrumental analysis methods used to determine trace amounts of selenium include flame atomic absorption spectrometry, graphite furnace atomic absorption spectrometry, hydride generation atomic absorption spectrometry, inductively coupled argon plasma optical emission spectrometry, inductively coupled plasma mass spectrometry, and neutron activation analysis.

Other methods of instrumental analysis include polarography, potentiometry, emission spectroscopy, x-ray diffraction, and x-ray fluorescence.

A widely used procedure for determining trace amounts of selenium involves separating selenium from solution by reduction to elemental selenium using tellurium (as a carrier) and hypophosphorous acid as reductant. The precipitated selenium, together with the carrier, are collected by filtration and the filtered solids examined directly in the wavelength-dispersive x-ray fluorescence spectrometer (70). Numerous spectrophotometric and other methods have been published for the determination of trace amounts of selenium (71–88).

Health and Safety Factors

Commercial elemental selenium along with the stable metallic selenides are considered relatively nontoxic. However, other selenium compounds which include the reactive selenides; the gaseous, volatile, and soluble compounds; and particularly hydrogen selenide, the halides, oxides, and the organics are highly toxic and must be handled with care (89). Selenium can enter the body through inhalation, ingestion, or absorption through the skin where it accumulates primarily in the liver and kidney. Symptoms of selenium poisoning include bronchial irritation, gastrointestinal distress, nasopharyngeal irritation, and garlic odor on the breath (90).

Selenium plays a dual role in a living organism, depending on the compound and the amount adsorbed. Controlled small doses of some compounds are used in medicine and as diet supplements, for example, ca 0.1 ppm of diet dry matter for livestock (see FEED ADDITIVES; MINERAL NUTRIENTS). Larger amounts can be toxic.

Contact with elemental selenium does not injure the skin. Selenium dioxide, however, upon contact with water, sweat, or tears, forms selenous acid, a severe skin irritant. Selenium oxyhalides are extremely vesicant and cause burns when in contact with human skin (91,92). Hydrogen selenide affects the mucous membranes of the upper respiratory tract and the eyes (93).

Industrial precautions include the common-sense measures of good housekeeping, proper ventilation, personal cleanliness, frequent changes of clothing, and provision of respirators (where needed), gloves, and either safety glasses or chemical goggles. Food should be consumed in a clean room separate from the handling area and after washing hands and face. Workers should change out of their work clothes and shower at the end of the working day. Selenium and its commercial products are not a fire hazard. The toxicity of seleniferous plants ingested by domestic animals and instances of poisoning among humans consuming the products of these animals and food (cereal grains) grown on highly seleniferous soils have been reported (19). Measures have been taken to cease grazing and cultivation on such soils. The aspects of selenium in human and animal biology and in medicine have been treated (14,68,94–96). Toxicity is discussed in References 97–102.

Uses

Electrical and Optical. The electrical and optical uses are based on the semiconducting and photoresponsive properties of selenium in its amorphous and crystalline trigonal forms. In the metastable amorphous form, the electrical conductivity is very low, making selenium almost an insulator, but when exposed to light the conductivity is greatly increased. This effect is exploited in xerography and in vidicons. The crystalline trigonal form has a higher electrical conductivity, which also increases under illumination. This conductivity increase, or photoconductivity, has its principal maximum at a wavelength of about 700 nm. The photoconductive effect in selenium was discovered by Smith in 1873 (103).

Trigonal selenium is a *p*-type semiconductor with an energy gap of 1.85 eV (104) and a work function of about 6 eV (105), which is the largest value reported for all the elements. Accordingly, a Schottky barrier should be created at the contact of selenium with any metal. This is consistent with the

observed nonsymmetric current–voltage characteristics for metal–Se contacts, particularly where the metal has a low work function. This property is used in rectifiers.

The photovoltaic effect, where an internal electromotive force is created, was discovered by Becquerel (106) in 1839 in an electrolyte with selenium but the effect in a metal–Se contact was first reported by Adams and Day (107) in 1876. The first practical photovoltaic cell was constructed by Uljanin (108) in 1888.

Rectifier. The selenium rectifier typically has the form Al–Ni–Se–CdBi as a layer structure. A base plate of steel or aluminum, sandblasted or etched, is plated with a film, about 1 μm thick, of bismuth or nickel to form a low resistance back contact. Next, a layer 50–60 μm thick of selenium doped with a halogen, is evaporated on the Ni- or Bi-covered heated plate. During or after this, the selenium is converted completely into the trigonal polycrystalline form by annealing at a temperature below the melting point of selenium (217°C). Then the selenium is coated with a counterelectrode of cadmium alloy, usually by spraying. Following this, the rectifier must be electrically formed, which is done by applying a reverse voltage continuously to the device for a period of hours or days. This process causes the reverse breakdown voltage to increase, without excessively increasing the forward resistance, thus improving the rectification ratio. The action has been ascribed to the formation of CdSe at the interface between the selenium and the counterelectrode. However, more recent work (109) indicates that the forming is more likely to be a partial sweep-out of acceptors in the selenium near the contact by ionic field drift.

Selenium rectifiers were extensively used in electrical power equipment from the 1940s to the 1960s. In this period, there were many variations in the processes used by the rectifier manufacturers. For example, it was found that the electrical characteristics were improved by the addition of some thallium to the selenium or the counterelectrode. Subsequent work showed that a counterelectrode of thallium alone gave even better rectification but the device suffered from rapid degradation owing to the chemical reactivity of thallium with the selenium. As of 1996, selenium rectifiers have been almost completely displaced by silicon diodes, which have a larger reverse breakdown voltage, smaller reverse leakage current, operate up to a higher temperature, and have more precise and controlled characteristics. Nevertheless, the selenium rectifier in single-plate form has a self-healing property not possessed by the silicon device. This property is demonstrated when breakdown by thermal runaway occurs at a hot-spot and localized melting takes place, producing a small crater with a wall of high resistance amorphous selenium, which avoids a short circuit. When silicon melts, it becomes metallic and the short circuit destroys the rectifier.

Photovoltaic Cell. The selenium photovoltaic cell is similar in basic structure to the rectifier, but instead of the opaque cadmium alloy, this cell has a transparent counterelectrode, usually of cadmium oxide deposited by reactive sputtering, on which an annular metal ring is deposited as the top contact. As in the rectifier, the selenium is in the form of a layer of polycrystalline trigonal material, usually doped with a halogen. The spectral response extends mainly over the wavelength range of 400 to 700 nm, with a maximum near 500 nm, which is similar to the wavelength sensitivity of the human eye. For this reason, the cell has been used as a photometer for measuring visible light levels. However, the device has been replaced to a large extent in this application by the CdS photoconductive cell. Under standard sunlight (AM1.5), the commercial selenium photovoltaic photometric cell yields an open-circuit voltage of 0.6 volt, a short-circuit current density of about 2 mA cm^2 , and a solar energy conversion efficiency below 1%. However, somewhat higher performance has been obtained in laboratory cells (110) (see PHOTODETECTORS; PHOTVOLTAIC CELLS).

Xerography. Photocopiers and laser printers, using the xerographic principle of operation (see ELECTROPHOTOGRAPHY), are a primary application for selenium and selenium-based alloys. A large number of such machines use selenium in the drum or belt photoconductor, which is the critical component for both the image generation and image development functions. The selenium is typically deposited on the metal belt or aluminum drum used in these machines, as a 50- μm vitreous layer, using a process which involves the thermal evaporation of high purity grades of selenium under vacuum. Early copiers used pure selenium coatings, but these have been almost totally replaced by alloys using additions of arsenic, chlorine, and tellurium in order to improve the lifetime of the drums, to modify their photoconductive or spectral response, and to improve their resistance to abrasion. As a result of these changes, some types of selenium-based photocopiers and printers can operate at very fast speeds and deliver hundreds of thousands of pages of output before the drum or belt needs to be replaced. Used selenium drums or belts are returned to the supplier for recycling. In the lower speed, lower cost, desktop digital copier, selenium has been largely replaced by organic-based photoconductors.

Selenium has also been used for xeroradiography. In this case the photoconductor is typically in the form of a flat plate and the image is formed by x-rays rather than by light. The original units of this type produced an x-ray image on paper, but these machines, which were primarily used in mammography, are nearly obsolete. Newer techniques for electronically developing the latent image formed on the reusable, selenium plate have been proposed (111,112). These could be the base for a new generation of digital radiographic systems, which would not require the use of traditional x-ray films.

Metallurgical. The metallurgical applications of selenium normally involve its use as a minor alloying additive to enhance the properties of both ferrous and nonferrous metals and alloys (see IRON; STEEL).

Ferrous Metals. In small concentrations, selenium decreases the surface tension of molten steel (qv) more than S, N, O, P, and C, but probably less than Te. In cast iron of low (0.02%) Mn content, selenium, in the presence of hydrogen, suppresses graphite nucleation and promotes carbide formation. However, at higher Mn levels graphite nucleation is favored. In steels, small amounts of Se act as a mild deoxidizer and selenium promotes fine-grained equiaxial crystals which minimize differences in directional properties, reduces hardenability, and is less sensitive to overheating and quench cracking. These properties result in improved steels for carburizing and hot-rolled applications. Selenium is also added to steel to counteract, on hot rolling, the elongation of globular MnS inclusions frequently encountered which lead to directional properties. It dissolves in the MnS, hardening the inclusion so that it will remain more oval during rolling. As little as 0.2% Se added to stainless steel casting alloys prevents the pinhole porosity ascribed to hydrogen.

The principle use of selenium in ferrous alloys is to improve machinability, ie, production rate and surface finish. Based on a machinability rating of 100 for AISI1112 steel, the machinability of low carbon steel 12L14, with the addition of 0.04–0.08% Se and 0.15–0.35% Pb, is increased 2.5 times. Similarly, the addition of a minimum of 0.15% Se to 302 and 430 stainless steels improves the machinability by 25 and 50%, respectively. The influence of selenium on improving the machinability of higher carbon grades of stainless steels is less pronounced. Even though sulfur also imparts free machining properties, selenium on a weight basis is preferred for better machinability and surface finish because selenium has a less deleterious effect on corrosion resistance, hot shortness, cold working, and strength properties of the finished product.

Selenium is added up to 0.1% to silicon steels (2–4% Si) used in transformer cores to enhance the development of the secondary recrystallization texture which, in turn, improves the magnetic characteristics. Selenium alloying additions to the melt may be made as elemental Se, nickel–selenium, or ferroselenium. The recovery depends on the melting practice and method of addition. Normally, it is in the range of 66%, but may be as high as 90%.

Copper and Copper Alloys. In liquid copper, selenium, owing to its surface activity, forms a monomolecular surface layer. The addition of selenium in the amount of 0.025–1% improves the machinability of copper. The microstructure of the alloy consists of primarily copper and a eutectic of copper plus cuprous selenide. The presence of the cuprous selenide phase acts as a chip breaker. Selenized copper has favorable brazing and soldering characteristics, although its weldability is degraded (see SOLDERS AND BRAZING FILLER METALS; WELDING).

Selenium has also been shown to act synergistically with bismuth to improve the machinability of brasses (113). The machining properties are similar to those of the leaded brasses used in plumbing applications. Environmental concerns arising from the leaching of lead brasses necessitates a replacement of the lead.

Lead and Lead Alloys. Selenium is reported to lower the surface tension of lead. The addition of 0.1% selenium and tellurium to solder improves its fluidity.

Selenium acts as a grain refiner in lead antimony alloys (114,115). The addition of 0.02% Se to a 2.5% antimonial lead alloy yields a sound casting having a fine-grain structure. Battery grids produced from this alloy permit the manufacture of low maintenance and maintenance-free lead–acid batteries with an insignificant loss of electrolyte and good performance stability.

Nickel–Iron and Cobalt–Iron Alloys. Selenium improves the machinability of Ni–Fe and Co–Fe alloys which are used for electrical applications. Neither sulfur nor tellurium are useful additives because these elements cause hot brittleness. The addition of 0.4–0.5% selenium promotes a columnar crystal structure on solidification, doubling the coercive force of cobalt–iron–titanium alloy permanent magnets produced with an equiaxial grain structure.

Chromium Plating. Sodium selenate or selenic acid are added to chromium-plating baths to improve corrosion protection from pitting,

blistering, and rusting, especially of plated automobile articles exposed to salt in the North American snow belt. The optimum concentration of the selenate ion is 0.015–0.018 g L in a bath containing 250 g L chromic acid and 2.5 g L sulfuric acid at a cathodic current density ca 21 A m² (225 A ft²) and 43–44°C. The selenate ion causes ca 600 uniformly distributed microcracks per linear centimeter in the plate. The network of fine intersecting cracks distributes the galvanic action over the surface and thus offers corrosion protection of the base metal. Such chromium plate has a dull luster with a decreased glare. A similar microporous, decorative, corrosion-protective plate has been developed (116) (see ELECTROPLATING).

Miscellaneous. Addition of selenium to magnesium and magnesium alloys is claimed to improve corrosion resistance by seawater. However, the technique of making this addition requires great care. Selenium dioxide has been added to baths used in etching, formation of decorative colors on galvanized iron and nonferrous metals and alloys, and corrosion-protective coatings in magnesium alloys. Inorganic and organic selenium compounds are brighteners in electroplating. Selenous acid improves the current efficiency in electrowinning of manganese. Diffusion coating of ferrous metals with selenium in a molten salt bath improves resistance to wear and to seizure of contact surfaces. In the manufacture of friction brake facings by powder metallurgy, the addition of selenium to copper oxide followed by reduction, mixing with nonmetallic particles, pressing, and sintering improves the adhesion of the metallic to the nonmetallic particles and increases the pressed green strength by 30^o %.

Glass and Ceramics. Selenium is used to decolorize the greenish tinge in glass caused by the presence of iron as an impurity. This application was introduced into the United States in 1915 (4). Until that time manganese dioxide imported from Russia was used in conjunction with arsenic trioxide. Selenium is usually used in the elemental form although other forms such as selenites of sodium, barium, or zinc and frits containing selenium may be used. The quantity required for a glass batch depends on the iron content of the sand used, preferably 0.02–0.05 wt % Fe₂O₃, and the selenium is added typically at a level of 10–30 g t of glass. Selenium retention efficiencies are generally around 20^o %. Utilization is being addressed by improving the form of the additive, modifying the furnace design, and improving the off-gas cleaning system (117).

Significant amounts of selenium are used to decolorize glass for container applications and to a lesser extent for table, lighting, and industrial glass applications. The flat plate glass industry uses selenium to produce gray to nearly black and bronze tinted heat-absorbing glasses for architectural and automotive applications. These glasses contain selenium in the range of 0.1–0.5 g kg in addition to measured amounts of cobalt and iron.

Selenium is combined with cadmium sulfide in various ratios to produce glasses colored in the range of orange to ruby red for the art and novelty glass industry. To produce the desired color, a glass batch containing the appropriate ratio of selenium and cadmium sulfide is melted under reducing conditions and rapidly cooled. The rapid cooling retains the color particles in solution and the glass takes on a faint yellow color. It is then subjected to a heat treatment called striking to develop the desired color. To produce a ruby red glass, which is influenced by melting conditions, Se, on the order of 0.5–1.0 kg, and CdS, 1.5–2.0 kg, are added to a 100-kg batch of sand. Selenium ruby glass exhibits very good transmission in the red part of the spectrum with a sharp cut-off of other colors and for this reason has found application in the United States since 1895 for signal light lenses, in addition to religious articles, vases, dinnerware, etc.

Selenium pink or rosaline glasses are obtained with the addition of 0.05–1.0^o % Se under oxidizing melting conditions. Selenium also imparts a yellow color to glass when oxides of antimony, arsenic, and bismuth are added. Addition of neodymium oxide and selenium produces purple glass and additions of cobalt or iron oxide with selenium yield black glasses. Selenium and silicon are used together to color glasses in conjunction with a third additive, eg, ferrous oxalate, to produce amber, bismuth oxide to produce topaz, and chromium oxide to produce emerald glass.

Selenium pigments are used in ceramic applications. These pigments may be encapsulated in zirconium silicate (118) which combines the brilliance of the pigments with high temperature and corrosion resistance under conditions of firing. Colored glazes of selenium are used in the ceramic industry for chinaware, porcelain and pottery, sanitary ware, and in printing inks for glass and other containers (see COLORANTS FOR CERAMICS).

Selenium chalcogenide glasses exhibit good infrared transmission properties. These are used as lenses (ZnSe, CdSe) in laser applications and have potential applications in fiber optics (qv) and in data storage and retrieval.

Pigments. Selenium combines with cadmium sulfide to produce cadmium sulfoselenide pigments with colors in the range of orange to maroon depending on the Cd–S–Se ratio. These are characterized by high temperature stability, brilliance of color, high chemical and light resistance, high opacity, and good dispersion in plastics and paints which are critical properties for use in ceramic, plastic, and glass applications. The olive-drab zinc–chromium–selenate pigments have corrosion-resistant properties (see CORROSION AND CORROSION CONTROL). A zinc sulfide-based pigment containing copper–indium–selenide sulfide exhibits an intensely golden yellow color (119) (see PIGMENTS, INORGANIC).

Rubber. The rubber industry consumes finely ground metallic selenium and Selenac (selenium diethyl dithiocarbamate, R. T. Vanderbilt). Both are used with natural rubber and styrene–butadiene rubber (SBR) to increase the rate of vulcanization and improve the aging and mechanical properties of sulfurless and low sulfur stocks. Selenac is also used as an accelerator in butyl rubber and as an activator for other types of accelerators, eg, thiazoles (see RUBBER CHEMICALS). Selenium compounds are useful as antioxidants (qv), uv stabilizers, (qv), bonding agents, carbon black activators, and polymerization additives. Selenac improves the adhesion of polyester fibers to rubber.

Lubricants. Selenium and its compounds are added to lubricating oils and greases for normal and extreme pressure service (see LUBRICATION AND LUBRICANTS). Dialkylselenides are oxidation inhibitors in lubricants. Barium, calcium, and zinc salts of selenic acids, RSe(O)OH, where R is an alkaryl radical with 10–40 carbon atoms, improve the detergent properties of lubricating oils. Dry, powdered lubricants containing diselenides of tungsten, niobium, tantalum, and molybdenum are manufactured for use at elevated temperatures in an ultrahigh vacuum. Under such conditions, they possess lower outgassing properties, lower coefficients of friction, and higher stability than graphite and the corresponding metal sulfides. Some self-lubricating iron alloys contain selenium.

Organic Chemistry and Pharmaceuticals. Selenium dioxide is an important oxidizing agent and catalyst in the synthesis of organic chemical and drug products (120) (see HORMONES; VITAMINS). It has been used in the manufacture of cortisone and nicotinic acid (niacin). Selenium and its compounds are useful in a wide variety of organic reactions, eg, oxidation, ammoxidation, reduction, hydrogenation, dehydrogenation, and polymer treatment. Dehydrogenation of hydroaromatic compounds >250° C with the evolution of hydrogen selenide has been valuable in elucidating the structure of some natural products. Selenium is often used as a catalyst in Kjeldahl determinations to hasten the digestion of nitrogenous materials. Some organoselenium compounds act as miticides. Others have neuroleptic and antidepressant properties. An antibiotic selenomycin is known. An organoselenium compound is claimed to possess psychotropic activity (121) (see PSYCHOPHARMACOLOGICAL AGENTS).

One area of research is the replacement of sulfur with selenium to enhance the potency of organic compounds in pharmaceutical applications. This has seldom been successful and often the toxicity is increased. There are some exceptions, eg, selenazofurin, phenylaminoethyl selenide, ebselen, and selenotifen (64). Selenazofurin is a cytotoxic compound having antitumor properties, phenylaminoethyl selenide is used to reduce hypertension, ebselen inhibits a variety of inflammatory and tissue damaging reactions, and selenotifen is an antiallergic agent.

Medicine and Nutrition. A stabilized buffered suspension of selenium sulfide has been marketed for many years as Selsun Blue (Abbott Laboratories) for control of seborrheic dermatitis of the scalp. A similar sulfur or selenium sulfide shampoo containing a metallic cation complex has been prepared (122). Topical application of selenium sulfide controls dermatitis, pruritis, and mange in dogs (see COSMETICS; VETERINARY DRUGS).

In 1956 selenium was identified (123) as an essential micronutrient in nutrition. In conjunction with vitamin E, selenium is effective in the prevention of muscular dystrophy in animals. Sodium selenite is administered to prevent exudative diathesis in chicks, a condition in which fluid leaks out of the tissues; white muscle disease in sheep; and infertility in ewes (see FEED ADDITIVES). Selenium lessens the incidence of pneumonia in lambs and of premature, weak, and stillborn calves; controls *hepatosis dietetica* in pigs; and decreases muscular inflammation in horses. White muscle disease, widespread in sheep and cattle of the selenium-deficient areas of New Zealand and the United States, is insignificant in high selenium soil areas. The supplementation of animal feeds with selenium was approved by the U.S. FDA in 1974 (see FEED ADDITIVES). Much of selenium's metabolic activity results from its involvement in the selenoprotein enzyme, glutathione peroxidase.

Selenium deficiency has been identified in humans in a broad area of China, stretching from the northeast to the southwest. Two syndromes are evident: Keshan disease, an endemic cardiomyopathy which affects children, and Kashin-Beck disease, an edemic osteoarthropathy (big-joint), also mainly involving children, which occurs in eastern Siberia and parts of China. As a result of these observations, a protective effect for selenium has been proposed, and various selenium-containing preparations have been applied to preserve animal and human health.

Inorganic and organic selenium compounds, including selenoproteins, have been used to treat selenium deficiency in humans and warm-blooded animals, control stress and high blood pressure in poultry and animals, combat aging, and improve the healing of surgically incised, lacerated, and burned

tissue (see ANTIAGING AGENTS; MEMORY ENHANCING DRUGS). Some organoselenium compounds are useful in immunoassays (qv). Tablets containing several micrograms of selenium as sodium selenite or as selenized yeast are sold over-the-counter in the United States and Europe.

Sufficient empirical data have been assembled to justify a broad research program enquiring into the protective effects of selenium against certain types of cancer, including colorectal and breast cancers (124). A massive field experiment was conducted in China in 1992 which showed that a treatment including selenium and some other antioxidants gave significant protection in an area of known high cancer incidence (125).

Miscellaneous. Sodium selenite is used in photography (qv) to produce prints with warm brown tones; sodium selenosulfate gives similar results. Derivatives of selenazole, benzoselenazole [273-91-6], and naphthoselenazole and selenium-containing dyes are used in supersensitive photographic emulsions. Selenocarbazides are used in photographic bleach fixing baths. Some photographic antifogging agents contain selenium.

Sodium selenate has been used on a small scale in commercial greenhouses, primarily for growing carnations and chrysanthemums. It is transformed by the plants into volatile selenides, which repel red spiders, mites, thrips, and aphids (see INSECT CONTROL TECHNOLOGY). Sodium selenite is not intended for crops which could ultimately be used as food for humans or domestic animals.

Sodium selenite has also been incorporated into styrene–butadiene rubber and used in a pellet form which results in the slow release of selenium into water. These pellets have been placed in lakes in Sweden which have fish contaminated with mercury owing to high levels of that element in the water. The selenium released by the pellets reacts with mercury to form insoluble, heavy mercury selenide which settles to the lake bottom and removes mercury from the ecosystem (126).

Selenium and selenium compounds are also used in electroless nickel-plating baths, delayed-action blasting caps, lithium batteries, xeroradiography, cyanine- and noncyanine-type dyes, thin-film field effect transistors (FET), thin-film lasers, and fire-resistant functional fluids in aeronautics (see ELECTROLESS PLATING).

The Selenium-Tellurium Development Association (STDA), located in Grimbergen, Belgium, and supported by the principal selenium producers, sponsors activities promoting the use of selenium, the development of new markets, and provides information in all areas where selenium is used. A newsletter is published semiannually. STDA technical efforts as of 1996 were being focused on the use of selenium as an additive to free machining plumbing brasses owing to the phase-out of leaded alloys in this application. When selenium is used in conjunction with bismuth, the elements act synergistically such that alloys with only 1–2% selenium and bismuth have close to the same machining properties as alloys containing the traditional 7% lead.

The use of high purity selenium in digital imaging is considered a potential market. Selenium-based plates can be used to replace conventional silver-based x-ray film, and in 1994 the Xerox Corporation introduced a selenium-based product, Verdefilm, for use in graphic arts applications.

Selenium is also used in thin-film photovoltaic cells (qv) which contain copper indium diselenide [12018-95-0], CuInSe₂. Use is quite small as of 1996. However, if the United States solar energy output with such cells were to increase by 100 MW/yr, this would require 8 t of selenium annually (see SOLAR ENERGY).

BIBLIOGRAPHY

"Selenium and Selenium Compounds" in *ECT* 1st ed., Vol. 12, pp. 145–163; "Selenium" in *ECT* 2nd ed., Vol. 17, pp. 809–833, by E. M. Elkin, Canadian Copper Refineries Ltd., and J. L. Margrave, University of Wisconsin; "Selenium and Selenium Compounds" in *ECT* 3rd ed., Vol. 20, pp. 575–601, by E. M. Elkin, Noranda Mines, Ltd.

1. K. W. Bagnall, *The Chemistry of Selenium, Tellurium, and Polonium*, 2nd ed., Elsevier Publishing Co., Amsterdam, the Netherlands, 1966.
2. M. Schmidt, W. Siebert, and K. W. Bagnall, *The Chemistry of Sulphur, Selenium, Tellurium, and Polonium*, Pergamon Press Ltd., Oxford, U.K., 1973.
3. N. V. Sidgwick, *The Chemical Elements and Their Compounds*, Oxford University Press, London, 1950, pp. 948–994.
4. R. A. Zingaro and W. C. Cooper, *Selenium*, Van Nostrand Reinhold Co., New York, 1974.
5. A. A. Kudryavtsev, *The Chemistry and Technology of Selenium and Tellurium*, Collets (Publishers) Ltd., London and Wellingborough, 1974.
6. *Gmelins Handbuch der Anorganischen Chemie*, 8th ed., Verlag Chemie, Weinheim Bergstrasse, System-Nummer 10, Teil A, Lfg.2, 1950, Lfg.3, Teil B, 1949.
7. J. N. Friend, ed., *Textbook of Inorganic Chemistry*, Griffin, London, Vol. 7, Part 2, 1931; Vol. 9, Part 4, 1937.
8. J. W. Mellor, *Comprehensive Treatise on Inorganic and Theoretical Chemistry*, Vol. 10, Longmans, Green & Co., New York, 1930, pp. 693–932.
9. D. M. Yost and H. Russel, Jr., *Systematic Inorganic Chemistry of the Fifth-and-Sixth-Group of Nonmetallic Elements*, Prentice Hall, Inc., Englewood Cliffs, N.J., 1944.
10. P. Pascal, *Nouveau Traité de Chimie Minérale*, Vol. XIII, Part 2, Masson & Cie, Paris, France, 1960, pp. 1651–1912.
11. K. J. Irgolic and M. V. Kudchadker, in Ref. 4, p. 408.
12. E. P. Painter, *Chem. Rev.* **28**, 179 (1941).
13. H. E. Gantner, in Ref. 4, p. 546.
14. D. L. Klayman and W. H. H. Gunther, eds., *Organic Selenium Compounds: Their Chemistry and Biology*, John Wiley & Sons, Inc., New York, 1973.
15. Y. Okamoto and W. H. H. Gunther, eds., *Ann. N.Y. Acad. Sci.* **192** (Apr. 17, 1972).
16. "2nd International Symposium on Organic Selenium and Tellurium Chemistry, Including Biochemistry," *Chem. Scr.* **8A** (1975).
17. N. D. Sindeeva, *Mineralogy and Types of Deposits of Selenium and Tellurium*, Interscience Publishers, a division of John Wiley & Sons, Inc., New York, 1964.
18. A. M. Lansche, *Selenium and Tellurium—A Materials Survey*, U.S. Bureau of Mines Information Circular 8340, Government Printing Office, Washington, D.C., 1967.
19. I. Rosenfeld and O. A. Beath, *Selenium—Geobotany, Biochemistry, Toxicity, and Nutrition*, Academic Press, Inc., New York, 1964.
20. B. Mason, *Principles of Geochemistry*, 3rd ed., John Wiley & Sons, Inc., New York, 1966.
21. S. P. Bapu, ed., *Trace Elements in Fuel, Advances in Chemistry*, Series 141, American Chemical Society, Washington, D.C., 1975.
22. J. E. Oldfield, *Selenium In Maps*, Bulletin, Selenium–Tellurium Development Association, Grimbergen, Belgium, Apr. 1995, pp. 1–7.
23. H. F. Mayland, L. F. James, K. E. Painter, and J. L. Scnderegger, *Selenium in Agriculture and the Environment*, Special Pub. 23, Soil Science Society and American Society of Agronomy, Madison, Wis., 1989, pp. 15–50.
24. S. M. Jasinski, *Proceedings of the Fifth International Symposium on Industrial Uses of Selenium and Tellurium, Brussels, Belgium, May 1994*, Selenium–Tellurium Development Association, Grimbergen, Belgium, pp. 13–20.
25. *Atlas of Electrochemical Equilibrium in Aqueous Solutions*, 2nd ed., Marcel Pourbaix, National Association of Corrosion Engineers, Houston, Tex.
26. W. A. Dutton, A. J. Van Den Steen, and N. J. Themelis, *Recovery of Selenium from Copper Anode Slimes*, paper no. A 71-54, TMS, Warrendale, Pa., 1971.
27. I. Fujimura and A. Katai, *Selenium Recovery from Copper Electrolysis Slimes at Mitsubishi Osaka Refinery*, paper no. 82-12, TMS, Warrendale, Pa.
28. J. E. Hoffmann, "Recovery of Selenium from Electrolytic Copper Refinery Slimes," in V. Kudryk, D. A. Corrigan, and W. W. Liang, eds., *Precious Metals: Mining, Extraction and Processing, V*, TMS, Warrendale, Pa., 1983.
29. **U.S. Pat. 3,249,399** (Mar. 29, 1966), J. E. Hoffmann, and D. C. Cusanelli (to Amax Inc.).
30. **U.S. Pat. 3,658,510** (Apr. 25, 1972), J. E. Hoffmann and R. Ernst (to Amax Inc.).
31. **U.S. Pat. 3,996,046** (Dec. 7, 1976), J. E. Hoffmann, P. J. Parker, and A. Sabo (to Amax Inc.).
32. D. H. Jennings, R. T. McAndrew, and E. S. Stratigakos, *A Hydrometallurgical Method for Recovering Selenium and Tellurium from Copper Refinery Slimes*, paper no. A 68-69, TMS, Warrendale, Pa., 1968.

33. P. H. Jennings and J. C. T. Farge, *Can. Min. Metall. Bull.*, 193–200 (Feb. 1966).
34. **U.S. Pat. 4,770,700**, B. J. Josef, W. Jorg, and H. Worz (to Austria Metall Ag At).
35. M. Lux, B. Mende, and S. Ziegengalb, *Neue Hutte* **23**(5) (May 78).
36. R. K. Manahan and F. Loewen, "Treatment of Anode Slimes at the INCO Copper Refinery," paper presented at *CIM Meeting*, Halifax, Nova Scotia, 1972.
37. **U.S. Pat. 4,047,939** (Sept. 13, 1977), B. Morrison (to Noranda Minerals).
38. N. Jordanov and L. Futekov, *Talanta* **13**(1), 163–168 (1966).
39. **Eur. Pat. 0049169** (Apr. 7, 1982), J. A. Thomas, N. C. Nissen, M. C. E. Bell, and A. Illis (to Inco Ltd.).
40. **U.S. Pat. 2,990,248** (June 27, 1961), L. E. Vaaler (to Diamond Alkali Inc.).
41. **Ger. Pat. 27336** (Jan. 28, 1964), R. Wagenmann, E. Scherz, and B. Puchta.
42. U. C. Woh, C. C. Change, W. L. Cheng, and I. S. Shou, *Hydrometallurgy* **1**(3), 381–392 (Jan. 1983).
43. T. Yanagida and N. Hosoda, *Treatment of Copper Electrolysis Slimes at Osaka Refinery*, paper no. A 75-42, TMS, Warrendale, Pa., 1975.
44. **U.S. Pat. 4,094,668** (June 13, 1978), J. C. Yannopoulos and B. M. Borham (to Newmont Exploration Ltd.).
45. Can. Pat. 1,003,190 (Jan. 11, 1977), S. T. Henriksson (to Boliden AB).
46. U.S. Pats. 2,189,480 (Feb. 6, 1940) and 2,196,380 (Apr. 9, 1940), C. W. Hewlett (to General Electric Co.); **U.S. Pat. 2,300,196** (Oct. 27, 1942), A. R. Bozarth (to Harshaw Chemical Co.); **U.S. Pat. 2,306,109** (Dec. 22, 1942), R. H. Long (to Harshaw Chemical Co.).
47. **U.S. Pat. 2,583,799** (Jan. 29, 1952), J. H. Schloen and L. V. Franchetto (to Canadian Copper Refiners Ltd.).
48. **U.S. Pat. 2,347,128** (Apr. 18, 1944), W. F. Russel (to R. T. Vanderbilt Co.); **U.S. Pat. 2,351,985** (June 20, 1944), J. Loeffler (to Alien Property Custodian).
49. **U.S. Pats. 1,382,920, 1,382,921, and 1,382,922** (June 28, 1921), V. Lenher, *J. Am. Chem. Soc.* **42**, 2498 (1920).
50. **U.S. Pat. 2,486,464** (Nov. 1, 1949), C. W. Clark and J. H. Schloen (to Canadian Copper Refiners Ltd.).
51. R. C. Brasted, *Comprehensive Inorganic Chemistry*, Vol. 8, *Sulfur, Selenium, Tellurium, Polonium, and Oxygen*, D. Van Nostrand Co., Inc., Princeton, N.J., 1961.
52. *Inorganic Syntheses*, McGraw-Hill Book Co., Inc., New York, published at intervals since 1939.
53. J. Brauer, ed., *Handbuch der Präparativen Anorganischen Chemie*, Thieme, Leipzig, Germany, 1925–1941.
54. P. J. Durant and B. Durrant, *Introduction to Advanced Inorganic Chemistry*, John Wiley & Sons, Inc., New York, 1962, p. 864.
55. D. M. Chizikov and V. P. Shchastlivyi, *Selenium and Selenides*, Collets (Publishers) Ltd., London and Wellingborough, 1968.
56. G. R. Waitkins and C. W. Clark, *Chem. Rev.* **36**, 235 (1945).
57. J. E. Cone, R. Martin Del Rio, J. N. Davis, and T. C. Stadtman, *Proc. Nat. Acad. Sci. USA* **73**, 2659–2663 (1976).
58. W. M. Ching, *Proc. Nat. Acad. Sci. USA* **81**, 3010–3013 (1984).
59. V. Krishnan and R. A. Zingaro, in Ref. 6, p. 337.
60. J. F. Alicino and J. A. Kowland, in Ref. 15, pp. 1049–1081.
61. L. Martillaro and M. Russo, in Ref. 14, p. 816.
62. D. Hausknecht and R. Ziskind, *Health Effects of Selenium*, NTIS PB-250568, Electric Power Research Institute, EPRI EA-571-V1, Palo Alto, Calif., Jan. 1976.
63. V. A. Potapov and S. V. Amosova, *Proceedings of the Fifth International Symposium on Industrial Uses of Selenium and Tellurium*, Brussels, Belgium, May 1994, Selenium–Tellurium Development Association, Grimbergen, Belgium, pp. 121–122.
64. M. J. Parnham, E. J. Dowling, C. Broquet, and P. Broquet, in Ref. 63, pp. 75–80.
65. T. E. Green and M. Turley, in I. M. Kolthoff and P. J. Elving, eds., *Treatise on Analytical Chemistry*, Vol. 7, Part 2, Interscience Publishers, a division of John Wiley & Sons, Inc., New York, 1961, pp. 137–205.
66. H. W. Furman, ed., *Standard Methods of Chemical Analysis*, 6th ed., Vol. 1, Part B.D., Van Nostrand Co., Inc., Princeton, N.J., 1962.
67. A. I. Vogel, *A Text-Book of Quantitative Inorganic Analysis*, 3rd ed., Longmans, Green & Co., London, 1962.
68. U.S. National Research Council, *Medical and Biologic Effects of Environmental Pollutants—Selenium*, U.S. National Academy of Sciences, Washington, D.C., 1976.
69. M. Yusuf, A. C. Sarki, S. B. Idris, G. A. Ayoko, and K. Singth, *Talanta*, **35**(6), 496–498 (1988).
70. B. A. T. Hoder, *Analyst* **114**, 919–922 (Aug. 1989).
71. H. Afsar, *Analyst* **114**, 1315–1318 (Oct. 1989).
72. K. N. Ramachandran, R. Kavershwar, and V. K. Gupta, *Talanta*, **40**(6), 781–784 (1993).
73. J. L. Bernal, M. J. Del Nozal, L. Deban, and F. J. Gomez, *Talanta*, **37**(9), 931–936 (1990).
74. A. Afkhami, A. Safavi, and A. Massoumi, *Talanta* **39**(8), 993–996 (1992).
75. D. L. Styris, L. J. Prell, and D. A. Redfeld, *Anal. Chem.* **63**, 508–517 (Mar. 1991).
76. E. H. Larsen, *Analyst*, **114**, 915–917 (Aug. 1989).
77. L. Shaopu, Z. Guangming, and H. Zhigui, *Talanta* **37**(7), 749–752 (1990).
78. J. Pettersson and A. Olin, *Talanta*, **38**(4), 413–417 (1991).
79. C. Ericzon, J. Pettersson, and A. Olin, *Talanta*, **37**(7), 725–730 (1990).
80. A. Kuldvere, *Analyst*, **114**, 125–131 (Feb. 1989).
81. D. Bakhtar, G. R. Bradford, and L. J. Lund, *Analyst*, **114**, 901–909 (Aug. 1989).
82. J. Bozic, D. Maskery, S. Maggs, and H. Susil, *Analyst*, **114**, 1401–1403 (Nov. 1989).
83. J. Bozic and co-workers, *Am. Lab.* **32F–32H**, 321 (Mar. 1994).
84. M. Borsier, *Spectrochim. Acta Rev.* **14**(1–2), 79–94 (1991).
85. H. Matusiewicz, *Spectroscopy*, **6**(2), 38–46 (Aug. 1990).
86. M. Tatro, *Spectroscopy*, **5**(3), 14–17 (1990).
87. J. R. Clark, *Anal. Chem.* **53**(1), 61–64 (Jan. 1981).
88. E. Mentasti, A. Nicolotti, V. Porta, and C. Sarzanini, *Analyst*, **114**, 1113–1117 (Sept. 1989).
89. *Documentation of the Threshold Limit Values and Biological Exposure Indices*, 6th ed., Vol. II, American Conference of Governmental Industrial Hygienists, Cincinnati, Ohio, 1991, pp. 1354–1360.
90. Clayton and Clayton, eds., *Patty's Industrial Hygiene and Toxicology*, 3rd rev. ed., John Wiley & Sons, Inc., New York, 1981.
91. J. J. Hostynek, R. S. Hinz, C. R. Lorence, P. Price, and R. H. Guy, *Crit. Rev. Toxicol.* **23**(2), 171–235 (1993).
92. H. C. Dudley, *Public Health Rep.* **53**, 94–98 (1938).
93. Ref. 89, pp. 784–785.
94. H. L. Cannon and H. W. Lakin, *Trace Metals in the Environment*, NTIS no. 274,428, U.S. Dept. of Commerce, Washington, D.C. and Research Triangle Park, N.C., Dec. 1976.
95. G. Combs, Jr., J. Spallholz, O. Levander, and J. Oldfield, *Selenium in Biology and Medicine: Third International Symposium, Beijing, The Peoples Republic of China, Parts A and B*, Van Nostrand Reinhold Co., New York, 1987.
96. W. Krause and P. Oehme, *Dtsch. Gesundheitswes* **34**(37) (1979).
97. J. Glover, O. Levander, J. Parizek, V. Vouk, in L. Friber and co-workers, eds., *Handbook of the Toxicology of Metals*, Elsevier North Holland

Biomedical Press, the Netherlands, pp. 555–577.

98. C. G. Wilbur, *Clin. Toxicol.* **17**(2), 171 (1980).

99. M. T. Lo and E. J. Sandi, *Environ. Pathol. Toxicol.* **4**, 193 (1980).

100. *Proceedings of the Fifth International Symposium on Industrial Uses of Selenium and Tellurium, Brussels, Belgium, May 1994*, Selenium–Tellurium Development Association, Grimbergen, Belgium.

101. O. A. Levander, *Selenium*, Environmental Health Criteria, Vol. 58, World Health Organization, Geneva, Switzerland, 1987.

102. Agency for Toxic Substances and Disease Registry (ATSDR), *Draft Toxicological Profile for Selenium*, U.S. Public Health Service, Atlanta, Ga., 1994.

103. W. Smith, *Nature* **7**, 303 (1873).

104. J. Stuke, in Ref. 4, p. 174.

105. C. H. Champness and A. Chan, *J. Appl. Phys.* **57**, 4823 (1985).

106. E. Becquerel, *Comptes Rendue Acad. Sci.* **9**, 561 (1839).

107. W. G. Adams and R. E. Day, *Proc. Roy. Soc.* **25**, 113 (1876).

108. W. Uljanin, *Annalen der Physik und Chemie* **27**, 241 (1888).

109. C. H. Champness and M. I. El-azab, *Solid State Electron.* **24**, 643 (1981).

110. C. H. Champness and C. H. Chan, *Solar Energy Materials and Solar Cells*, to be published.

111. **U.S. Pat. 4,961,209** (Oct. 2, 1990), H. J. Davis, N. Araj, and J. A. Rowlands (to Noranda Inc.).

112. D. M. Korn, S. J. Johnson, S. L. Nelson, and R. J. Ziegler, *J. Appl. Photo Eng.* **4**, 178 (1978).

113. M. G. King, D. Hayduk, and A. Mirza, in Ref. 63, pp. 51–54.

114. B. E. Kallup, *Proceedings of the Third International Symposium on Industrial Uses of Selenium and Tellurium, Saltsjybaden, Stockholm. Sweden, Oct. 1984*, Selenium–Tellurium Development Association, Grimbergen, Belgium, p. 109.

115. **U.S. Pat. 3,801,310** (Apr. 2, 1974), S. C. Nijhawan (to Varta AG).

116. **U.S. Pat. 4,007,099** (Feb. 8, 1977), S. H. L. Wu (to Harshaw Chemical Co.).

117. H. Möller-Simon and H. Barklage-Hilgefort, in Ref. 63, p. 273.

118. P. Kleinschmit, in Ref. 114, pp. 246–251.

119. Ger. Offen. 3,005,221 (Aug. 21, 1980), G. P. Kinstle (Ferro Corp.).

120. V. Kollonitsch and C. H. Kline, *Ind. Eng. Chem.* **55**(12), 18 (1963).

121. **USSR Pat. 551,327** (May 23, 1977), I. N. Azarbaev and co-workers (to Institute of Chemical Sciences).

122. **U.S. Pat. 4,089,945** (May 16, 1978), R. E. Brinkman and R. L. Vogenthaler (to The Procter & Gamble Co.).

123. K. Schwarz and C. M. Foltz, *J. Am. Chem. Soc.* **79**, 3292–3293 (1957).

124. B. S. Reddy, A. Rivenson, N. Kulkarni, P. Upadhyay, and K. El-Bayoumy, *Cancer Res.* **52**, 5635–5640 (1992).

125. W. J. Blot and co-workers, *J. Nat. Cancer Inst.* **85**, 1483–1492 (1993).

126. K. Paulsson and K. Lundbergh, in Ref. 63, pp. 287–290.

J. E. Hoffmann
Jan Reimers and Associates USA Inc.

M. G. King
Selenium–Tellurium Development Association

SEMICONDUCTORS

Silicon-based,
Amorphous,
Compound,

SILICON-BASED SEMICONDUCTORS

With the invention of the transistor in 1947 semiconductors changed the basis of electronics from vacuum tubes to semiconductor devices. Some of the history of the development of transistor technology has been described (1–12). The invention of the integrated circuit in 1958 revolutionized electronic systems. The reason for this can be seen by considering the Intel 8086 (1978), the 16-bit microprocessor used in the first IBM PC. This chip used NMOS technology (MOS = metal oxide semiconductor) with 10-μ feature lengths and contained 20,000 transistors in a 25 mm² chip. Compare this with the Whirlwind computer (1950), the first large-scale, real-time control system, which contained 15,000 vacuum tubes and occupied a 2500 ft² room. Thus, a single semiconductor chip contained the complexity of a large system. As important, the failure rate of the Intel 8080 with 4500 transistors, 220 in 10⁹ hours, was comparable to the failure rates of reliable, single transistors. Remarkably, this revolution in circuit integration has continued. Microprocessor chips already contain more than three million transistors and, according to the Semiconductor Industry Association's (SIA) National Technology Roadmap for Semiconductors, transistor densities are projected to rise to 90 million transistors in 1 cm² by 2010 (13).

Accompanying this change in the scale of integration has been a shift in the size and composition of the semiconductor industry. The silicon-based semiconductor industry is the basis for the technology sector, which includes computers, computer peripherals, and telecommunications, that has become the largest of the U.S. economy (14). Unfortunately, there is a shortage of reliable data for this sector because the system of U.S. economic statistics was developed for the traditional manufacturing sector shortly after World War II. As a result, statistics have to be obtained from a variety of private sources. According to one estimate, U.S. semiconductor sales grew from \$0.5 billion in 1961 to \$2.4 billion in 1976 (15). This was part of a \$5.4 billion world market which rose to \$102 billion by 1994 (16). MOS DRAMs (DRAM = dynamic random access memory) were 38% of the \$88 billion market for integrated circuits (ICs) (qv), and only \$14 billion were spent on discrete, single-transistor semiconductor devices (16). This has been a dynamic market, rising to \$144 billion in 1995, with integrated circuits (\$126 billion) growing more rapidly than discrete devices (\$18 billion) (17). In 1996 the semiconductor market is predominantly MOS digital integrated circuits.

This is the present context for silicon-based semiconductor technology. To a great extent, it has been driven by advances in MOS DRAM technology. Newer developments have been the result of steady advances in MOS manufacturing technology rather than the invention of new semiconductor devices. In 1979, increasing integration was predicted (18), with the number of MOSFETs per die quadrupling every three years (FET = field-effective transistor). This was a product of shrinking minimum feature size (~70% reduction 3 yrs) and increasing die area (~50% increase 3 yr). The reduction in minimum feature size has the greatest effect, corresponding to the product of the reductions in the horizontal and vertical directions. The consequences of this trend can be seen in Figure 1 (19) which compares the rates of technological change in the computer and industrial revolutions. If technological change is measured in DRAM memory bits or transistors for the microelectronics revolution and in the horsepower of steam engines for the industrial revolution, the result is exponential growth in both cases, which corresponds to straight lines on a semilogarithmic plot. Although the amount of technological change is comparable for both graphs, the industrial revolution took a hundred years to accomplish what the microelectronics revolution has accomplished in 10 years. Further, the microelectronics revolution is continuing to expand, with 64 Mb DRAMs already developed and 64 Gb DRAMs predicted for 2010 (13).

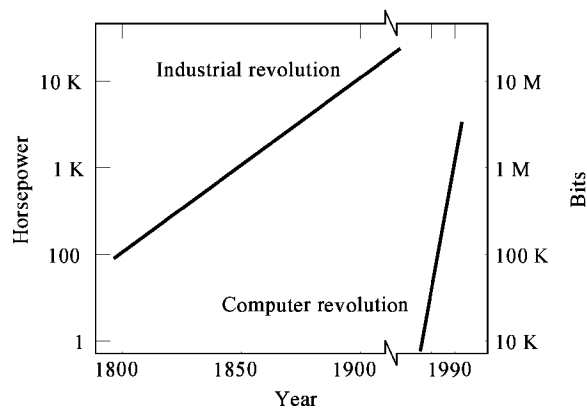


Fig. 1. The pace of technological change in both the microelectronics-based computer revolution, as characterized by the number of bits in a DRAM, and the industrial revolution, as characterized by the horsepower developed by a steam engine (19).

Courtesy of IEEE

As the silicon semiconductor field has matured, its focus has shifted from new devices to the manufacture (fabrication) of well-understood devices, notably MOSFETs, in ever-smaller dimensions, and "each new technology generation poses materials and process riddles different from those solved in previous generations. There are new types of contamination to worry about, and new sources of defects when the scale is shrunk" (20). An example of this has been the steady increases required in the extraordinary chemical purity necessary for semiconductor manufacture. The behavior of semiconductor devices is determined by changes in impurity concentrations on the order of 10¹⁷ cm⁻³ or 2 parts per million atoms (ppma), because the number density of silicon is 5 × 10²² cm⁻³. The starting material for semiconductor manufacturing is a single crystal of silicon, 200 mm (8 in.) in diameter and ~650-μm thick, with a global flatness variation across the wafer of <3 μm (21). Chemical perfection of these giant single crystals is essential (22). The largest impurity in these crystals, oxygen, is kept around 20–35 ppma. In this range, the oxygen improves the mechanical strength of the wafer and promotes intrinsic gettering (23), in which the oxygen is used to draw fast diffusing metal contaminants into the bulk of the wafer and away from the surface on which active devices are formed. Specific metallic contaminants, particularly heavy metals such as Fe, Ni, Cu, and Au, need to be kept below 0.01 ppba because their presence can degrade device performance by reducing minority carrier lifetimes (21).

Perfection especially is required on the silicon surface. A {100} surface of silicon contains 6.8 × 10¹⁴ atoms cm⁻². Surface defect densities must be less than one part in 10⁵–10⁶ defects cm⁻² for satisfactory MOSFET operation. In fact, the discovery of the original point contact transistor was only possible because the native oxide on single-crystal germanium has surface defect densities less than one part in 10⁴. Good silicon devices required the discovery (10) that the thermal oxidation of silicon could produce an excellent Si-SiO₂ interface.

The high purity required of silicon and the small size of semiconductor devices place stringent limits on the chemicals used in processing silicon. By far the most important chemical is water which is used extensively to dilute etchants and clean wafers. Pure water (24) is required to have fewer than 0.025

ppba of Na⁺ and Cl⁻, 0.002 ppba of heavy metals such as Fe, and <10 ppb of total oxidizable (organic) carbon (TOC), eg, bacteria. In addition, pure water should have less than one particulate per liter greater than a half micrometer in size. Water this pure is, in fact, an active chemical which absorbs, dissolves, or reacts with all kinds of materials to create contamination. Similar requirements are placed on other process chemicals.

Semiconductor Materials Theory

Silicon is a Group 14 (IV) element of the Periodic Table. This column includes C, Si, Ge, Sn, and Pb and displays a remarkable transition from insulating to metallic behavior with increasing atomic weight. Carbon, in the form of diamond, is a transparent insulator, whereas tin and lead are metals; in fact, they are superconductors. Silicon and germanium are semiconductors, ie, they look metallic, so that a polished silicon wafer is a reasonable gray-toned mirror, but they conduct poorly. Traditionally, semiconductors have been defined as materials whose resistance rises with decreasing temperature, unlike metals whose resistance falls.

Diamond, silicon, and germanium all crystallize in the diamond crystal structure with atoms tetrahedrally bonded to each other. This crystal structure can be described as two interpenetrating face-centered cubic (fcc) lattices (25). As one moves down the column of Group IV, the bonds lengthen and weaken, lowering the melting and boiling points of these materials. Table 1 indicates some of the key properties of Group IV semiconductors.

Table 1. Important Properties^a of Group IV Semiconductors

Material	Cubic lattice constant, pm	Band gap, eV	Intrinsic carrier concentration, cm ⁻³	Relative dielectric constant, ε _i	Mobility, cm ² (Vs)	
					Electrons	Holes
C (diamond)	356.683	5.47	7.6 × 10 ⁻²⁶	5.7	1800	1200
6H-SiC	^b	3.0	1.6 × 10 ⁻⁶	9.66		
SiC					380	75
Si	543.095	1.12	1.45 × 10 ¹⁰	11.9	1500	450
Ge	564.613	0.66	2.4 × 10 ¹³	16.0	3900	1900

^a At 300 K.
^b α-SiC crystallizes in the wurtzite lattice with *a* = 308.6 pm and *c* = 1511.7 pm (values from Refs. 26 and 27).

The electrical behavior of many materials can be explained elegantly by band theory (28). When atoms are packed together their energy levels split into bands of closely spaced energy levels. Some of these bands overlap and some are separated by energy gaps. Whether a material is an insulator or conductor depends on whether a band is completely or partially filled. In the case of silicon, the atom contains two electrons which fill the 3*s* level and two electrons in the six states of the 3*p* level. As the atoms are brought together in the diamond configuration, these eight states per atom hybridize and split, forming two bands of energy levels separated by an energy gap. At 0 K, the lower valence band with four bonding states per atom is filled and the upper conduction band, which also has four antibonding states per atom, is empty. Consequently, diamond, silicon, and germanium are essentially insulators with an empty conduction band separated from a filled valence band by an energy gap, *E_g*, of the size indicated in Table 1. At higher temperatures, some electrons are excited to the conduction band, leaving behind empty states in the valence band. Applying pressure decreases the lattice spacing and increases *E_g*; thermal expansion decreases *E_g*. These effects can be exploited in pressure- and temperature-sensitive transducers (29).

Near a conduction band minimum the energy of electrons depends on the momentum in the crystal. Thus, carriers behave like free electrons whose effective mass differs from the free electron mass. Their energy is given by equation 1, where *E_c* is the energy of the conduction band minimum, *p* is the

$$E = E_c + \frac{p^2}{2m^*}$$

(1)

effective momentum of the electrons in the crystal, and *m^{*}* is the effective mass. Because the crystal momentum *p* is related to the wave vector *k* = 2π / λ (by de Broglie's relation, *p* = ħ*k* = 2πħ / λ, where ħ is Planck's constant and ħ = 2π / 0.658 × 10⁻¹⁵ eVs = 0.685 feVs), the energy bands are often drawn as *E* vs *k* diagrams. The nearly filled states at the top of the valence band behave in conduction as positive charge carriers (holes) with a different effective mass. The density of states is the number of distinct states or *k*-values available for occupation within a given energy interval. There are two states for each unique *k*-value, corresponding to the two possibilities for electron spin. A low effective mass means that the *E* vs *k* diagram has a steep slope and a correspondingly low density of states. Figure 2a shows the *E* vs *k* diagram for an idealized semiconductor in which the electrons and holes have nearly equal mass. The solid lines in Figure 2a correspond to available states for electrons with quantized *k*-values. If the available states are counted as a function of energy, the density (per unit volume) of states (DOS) for electrons is given, as shown in Figure 2b.

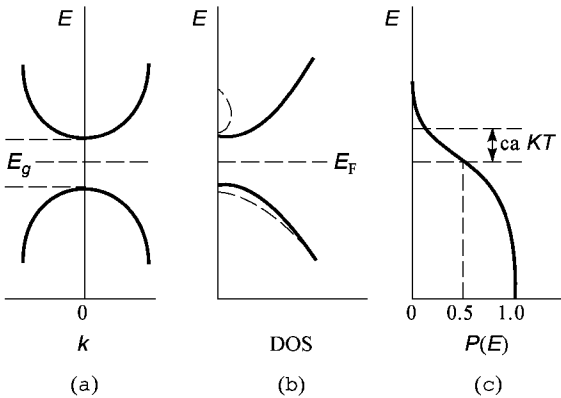


Fig. 2. (a) Energy, *E*, versus wave vector, *k*, for free particle-like conduction band and valence band electrons; (b) the corresponding density of available electron states, DOS, where *E_F* is Fermi energy; (c) the Fermi-Dirac distribution, ie, the probability *P*(*E*) that a state is occupied, where *K* is the Boltzmann constant and *T* is absolute temperature in Kelvin. The tails of this distribution are exponential. The product of *P*(*E*) and DOS yields the energy distribution of electrons shown by () in (b).

This simple picture is complicated because the dependence of energy on momentum may be anisotropic and the energy bands may have several minima or maxima. This is the case for silicon whose minimum conduction band energy surfaces are displaced spheroids along the <100> axes (30). This can be described by introducing longitudinal and transverse electron effective masses, *m_l* = 0.92 *m* and *m_t* = 0.19 *m*, respectively. An appropriate average gives the density of states effective mass for electrons at 4 K, *m^{*}* = 1.06 *m* (30).

The energy band theory helps to explain the optical properties of materials. Photons must have energies greater than the energy gap to excite electrons from the valence to the conduction band. Visible photon energies are in the red-to-blue energy range of 1.7 to 3 eV. Thus, insulators such as diamond are frequently transparent ($E_g > 3\text{ eV}$) and metals are opaque. Silicon with an energy gap of 1.12 eV is opaque to visible light but transparent in the near-infrared for wavelengths $\lambda(\mu\text{m}) > 1.24\text{ }E_g(\text{eV}) = 1.1\text{ }\mu\text{m}$. This transparency is important for applications such as solar cells which depend on the efficient absorption of light. Normally transparent crystals such as sapphire can be colored by the presence of impurities which introduce energy levels into the energy gap and alter the absorption characteristics.

Diamond, silicon, and germanium have an indirect band gap, ie, the conduction band minimum has a different momentum than the valence band maximum. Compound semiconductors, such as GaAs, generally have a direct band gap in which these momenta match. In direct band gap semiconductors an electron dropping from the conduction band minimum to fill an empty state at the valence band maximum can conserve energy and momentum by emitting a photon, which has very little momentum. The same transition in an indirect band gap semiconductor also requires emitting a quantum of lattice vibration (phonon) to conserve momentum. Because this three-body process is much less likely, optical emission is much less efficient than for direct semiconductors. Thus, the increasing interest in integrated optoelectronic devices requires a difficult marriage of silicon and compound semiconductor technology.

Remarkably, although band structure is a quantum mechanical property, once electrons and holes are introduced, their behavior generally can be described classically even for deep submicrometer geometries. Some allowance for band structure may have to be made by choosing different values of effective mass for different applications. For example, different effective masses are used in the density of states and conductivity (26).

Semiconductor Statistics

Intrinsic Semiconductors. For semiconductors in thermal equilibrium, $\langle N(E) \rangle$, the average number of electrons occupying a state with energy E is governed by the Fermi-Dirac distribution. Because, by the Pauli exclusion principle, at most one electron (fermion) can occupy a state, this average number is also the probability, $P(E)$, that this state is occupied (see Fig. 2c). In equation 2, K

$$\langle N(E) \rangle = P(E) = \frac{1}{1 + \exp(E - E_F / KT)}$$

(2)

is the Boltzmann constant and T is the absolute temperature in Kelvin. E_F , the Fermi energy, normalizes the distribution $P(E)$ at all temperatures. At low temperatures the Fermi-Dirac distribution is rectangular, with $P(E) \approx 0$ for $E > E_F$. However, at normal operating temperatures for a semiconductor device, $|E - E_F| > KT$, so the exponential in the denominator of equation 2 dominates and $P(E)$ is well approximated by a classical, exponential, Maxwell-Boltzmann distribution.

In an undoped, intrinsic semiconductor the equilibrium concentrations of electrons, n , and holes, p , are described by a lever rule derived from the law of mass action (eq. 3):

$$np = n_i^2 = N_c N_v \exp(-E_g / KT)$$

(3)

In an intrinsic semiconductor, charge conservation gives $n = p = n_i$, where n_i is the intrinsic carrier concentration as shown in Table 1. N_c and N_v are the effective densities of states per unit volume for the conduction and valence bands. In terms of these densities of states, n and p are given in equations 4 and 5,

$$n = N_c \exp[-(E_c - E_F) / KT]$$

(4)

$$p = N_v \exp[-(E_F - E_v) / KT]$$

(5)

where E_c and E_v are the energies of the conduction and valence band edges, $E_g = E_c - E_v$, and E_F is the Fermi level. Electrons and holes obey Fermi-Dirac statistics. The above formulas are the result of integrating, over all energies, the product of the density of states and the probability for finding a carrier at a particular energy. These formulas assume that the Fermi level is located in the energy gap, E_g , away from the conduction and valence band edges so that the statistics obey the classical Maxwell-Boltzmann limit. N_c and N_v are functions of temperature and are proportional to $T^{3/2}$.

$N_{c,v} = 2M_{c,v} (2\pi m_{c,v}^* KT / h^2)^{3/2}$ where $M_{c,v}$ the number of equivalent minima or maxima in the conduction and valence bands, respectively, and $m_{c,v}^*$ the density of states effective masses of electrons and holes.

Extrinsic Semiconductors. The most common impurities or dopants introduced into the silicon crystal are Periodic Table Group 15 (V) and Group 13 (III) elements. These dopants are substitutional, ie, they replace silicon atoms in the crystal lattice. Group 15 elements contain five valence electrons, four of which bond covalently with the surrounding silicon atoms. These impurities are called donors because their fifth electron is weakly bound and can be donated to the conduction band by thermal excitation. This weak bonding is approximately described by a hydrogen atom model in which the energy levels of the hydrogen atom are adjusted for the effective mass of the electron and the relative dielectric constant of the semiconductor, $\epsilon_r = 11.9$ for silicon. In silicon, As, P, and Sb have shallow donor energy levels, E_d , that are 0.045, 0.044, and 0.039 eV below E_c , respectively. A similar argument can be made for Group 13 impurities. In silicon, B, Al, Ga, and In have shallow acceptor levels, E_a , which are 0.045, 0.057, 0.065, and 0.16 eV above E_v , respectively. In practice, the most common donors are phosphorus and arsenic; the most common acceptor is boron.

The variations in E_d and E_a and the much larger value for In show the limitations of a simple hydrogen atom model. Other elements, particularly transition metals, tend to introduce several deep levels in the energy gap. For example, gold introduces a donor level 0.54 eV below E_c and an acceptor level 0.35 eV above E_v in silicon. Because such impurities are effective aids to the recombination of electrons and holes, they limit carrier lifetime.

The carrier concentrations in doped or extrinsic semiconductors to which donor or acceptor atoms have been added can be determined by considering the chemical kinetics or mass action of reactions between electrons and donor ions or between holes and acceptor ions. The condition for electrical neutrality is given by equation 6. When the predominant dopants are donors, the semiconductor is

$$n + N_a = N_d + p$$

(6)

***n*-type.** In *n*-type semiconductors the majority carriers are electrons and the minority carriers are holes so that $N_d > N_a$ and $n \approx N_d$. By the lever rule, $p \approx n^2 / N_d < n$, so the hole concentration is greatly reduced by the introduction of donors. Similarly, for *p*-type semiconductors, $N_a > N_d$ and $p \approx N_a$.

For lightly doped *n*-type semiconductors at normal operating temperatures there is complete donor dissociation (donor saturation). $N_d \approx N_d$, corresponding to E_F well below E_d . In the presence of acceptors, $n - N_d + N_a \approx N_d - N_a$. This frequently occurs because semiconductor devices are customarily made by the diffusion or implantation of excess donors into a *p*-type substrate or vice versa.

As the temperature increases, the intrinsic carrier concentration rises exponentially so that at some point $n \approx p \approx n_i$, rather than $n \approx N_d$. Likewise, as the temperature falls, $E_a - E_v$ becomes large compared with KT and carrier freeze-out occurs below about 100 K so that $n \approx N_d < N_d$. All of this assumes that $n < N_c$ or $p < N_v$, corresponding to E_F in the energy gap. When enough donor atoms are added so that $n > 10 N_c$ the Fermi level moves

into the conduction band, $E_F > E_c$, and the semiconductor is said to be degenerately doped or n^+ . Now Fermi-Dirac statistics apply and the degenerately doped semiconductor behaves like a metal. In particular, carriers do not freeze out at cryogenic temperatures. Similar comments apply to degenerately doped p^+ semiconductors.

The equilibrium lever relation, $np = n_i^2$, can be regarded from a chemical kinetics perspective as the result of a balance between the generation and recombination of electrons and holes (21). In extrinsic semiconductors recombination is assisted by chemical defects, such as transition metals, which introduce new energy levels in the energy gap. The recombination rate in extrinsic semiconductors is limited by the lifetime of minority carriers which, according to the equilibrium lever relation, have much lower concentrations than majority carriers. Thus, for a p -type semiconductor where electrons are the minority carrier, the recombination rate is $\Delta n / \tau_n$. $\Delta n = n - n_0$ is the increase of the electron concentration over its value in thermal equilibrium, n_0 , and τ_n is the minority carrier lifetime. This assumes low level injection where Δn is much smaller than p_0 , the equilibrium majority carrier concentration.

Recombination and generation rates generally are greatly accelerated by the presence of deep levels in the band gap because the required energy jumps are smaller. Deep levels act as traps which capture and emit carriers from the conduction or valence bands. To be effective as a recombination center a trap must capture an electron and subsequently capture a hole before the electron is emitted to the conduction band, or vice versa. Levels near midgap, such as gold provides in silicon, are most effective in increasing the recombination rate. This greatly decreases the minority carrier lifetime. Surface and interface states can also act as recombination centers, drastically reducing minority carrier lifetimes.

Noise. So far, as indicated at the beginning of this section on semiconductor statistics, equilibrium statistics have been considered. Actually, there are fluctuations about equilibrium values, $\Delta N = N - \langle N \rangle$. For electrons, the mean-square fluctuation is given by $\langle (\Delta N)^2 \rangle = \langle N \rangle (1 - \langle N \rangle)$ where $\langle N(E) \rangle$ is the Fermi-Dirac distribution. This mean-square fluctuation has a maximum of one-fourth when $E = E_F$. These statistical fluctuations act as electrical noise and limit minimum signal levels.

Semiconductor devices are affected by three kinds of noise. Thermal or Johnson noise is a consequence of the equilibrium between a resistance and its surrounding radiation field. It results in a mean-square noise voltage which is proportional to resistance and temperature. Shot noise, which is the principal noise component in most semiconductor devices, is caused by the random passage of individual electrons through a semiconductor junction. Thermal and shot noise are both called white noise since their noise power is frequency-independent at low and intermediate frequencies. This is unlike flicker or $1/f$ noise which is most troublesome at lower frequencies because its noise power is approximately proportional to $1/f$. In MOSFETs there is a strong correlation between $1/f$ noise and the charging and discharging of surface states or traps. Nevertheless, the universal nature of $1/f$ noise in various materials and at phase transitions is not well understood.

Semiconductor Transport

Charge carriers in a semiconductor are always in random thermal motion with an average thermal speed, v_{th} , given by the equipartition relation of classical thermodynamics as $m^* v_{th}^2 / 2 = 3KT / 2$. As a result of this random thermal motion, carriers diffuse from regions of higher concentration. Applying an electric field superposes a drift of carriers on this random thermal motion. Carriers are accelerated by the electric field but lose momentum to collisions with impurities or phonons, ie, quantized lattice vibrations. This results in a drift speed, v_d , which is proportional to the electric field; $v_d = \mu_e E$ where E is the electric field in volts per cm and μ_e is the electron's mobility in units of cm^2/Vs .

The electron current density J_e has units of A/cm^2 and in a semiconductor results from drift and diffusion. In the absence of concentration gradients, equation 7 reduces to Ohm's law, $J_e = nq\mu_e E = \sigma E$ where σ is the conductivity,

$$J_e = nq\mu_e E + qD_e \frac{\partial n(x)}{\partial x} = \mu_e n \partial E_F / \partial x$$

(7)

q is the magnitude of the charge of an electron, and D_e is the electron's diffusion constant. The Einstein relation relates μ_e and D_e by $D_e = KT\mu_e / q$. A similar equation can be written for hole currents, J_p , if $+qD_e$ is replaced by $-qD_h$. High mobilities generally lead to high currents. In practice, $\mu_h < \mu_e$ for most semiconductors.

Ohm's law assumes that the drift speed of electrons in an electric field, $v_d = \mu_e E$, is small compared to their average speed, v_T , in a Maxwell-Boltzmann distribution. At high electric fields, $E > 10 \text{ kV}/\text{cm}$, v_d no longer increases with electric field and approaches a limiting saturation speed, v_s , determined primarily by optical phonon emission. Figure 3 shows the variation of drift speed with electric field for electrons and holes in various semiconductors.

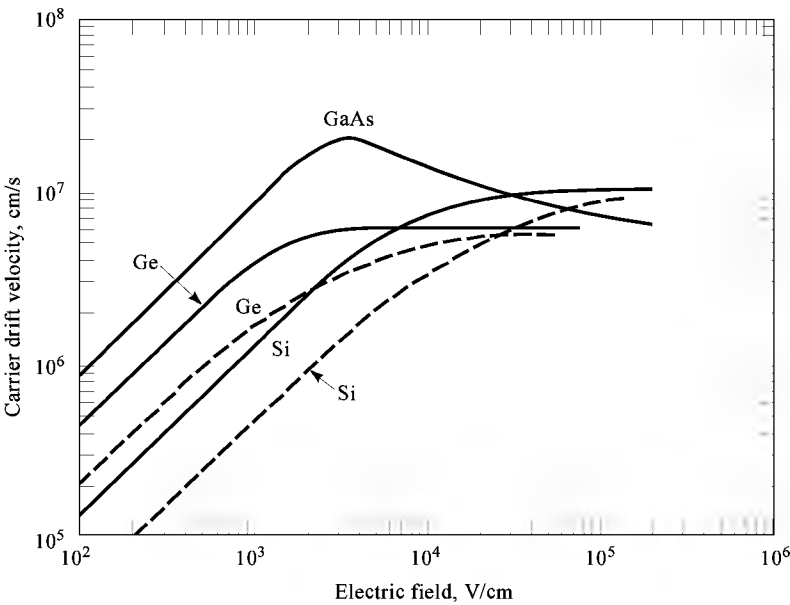


Fig. 3. Electron (—) and hole (---) velocities versus electric field for high purity silicon, Si; germanium, Ge; and gallium arsenide, GaAs, at 300 K (26).

At still higher fields carriers can acquire enough energy from motion in an electric field to create electron-hole pairs by impact ionization. For silicon the electron ionization rate, which is the number of pairs generated per cm of electron travel, depends exponentially on electric field. It is about $2 \times 10^3 \text{ cm}^{-1}$ for a $50 \text{ kV}/\text{cm}$ field at 300 K. The electric field causes electrons and holes so created to travel in opposite directions. They may create other electron-hole pairs causing positive feedback, which leads to avalanche breakdown at sufficiently high fields.

MOS Capacitance

So far, the fundamental factors which affect the current a device can carry have been considered. In practice, the speed of digital (and analog) devices is determined by the time it takes to charge or discharge a capacitor, C , where $\Delta Q = C \Delta V = I \Delta t$. Thus, from this general relationship, low values of Δt require some combination of low C , low ΔV , or large I . Different device technologies and circuit techniques achieve this in different ways, but it can be seen that an understanding of device and circuit capacitance is essential to an assessment of device performance.

This is particularly true for the MOSFET where conduction is controlled by a metal oxide semiconductor (MOS) capacitor. When a voltage is applied to the metal plate of a MOS capacitor most of the voltage drop is across the oxide but some drop is across the semiconductor. If the semiconductor is p -type, a positive voltage applied to the metal plate induces negative charges in the semiconductor. Initially, this causes holes to move away from the surface of the semiconductor, leaving behind a carrier depletion region containing immobile, ionized acceptors, n_a . Figure 4 shows the charge distributions and band bending in a MOS capacitor as the applied voltage is changed. As the voltage is increased, mobile electrons are drawn to the surface. Strong inversion is said to occur when the density of electrons at the surface equals the original density of holes. Strong inversion is generally used to define the threshold voltage for conduction in MOSFETs.

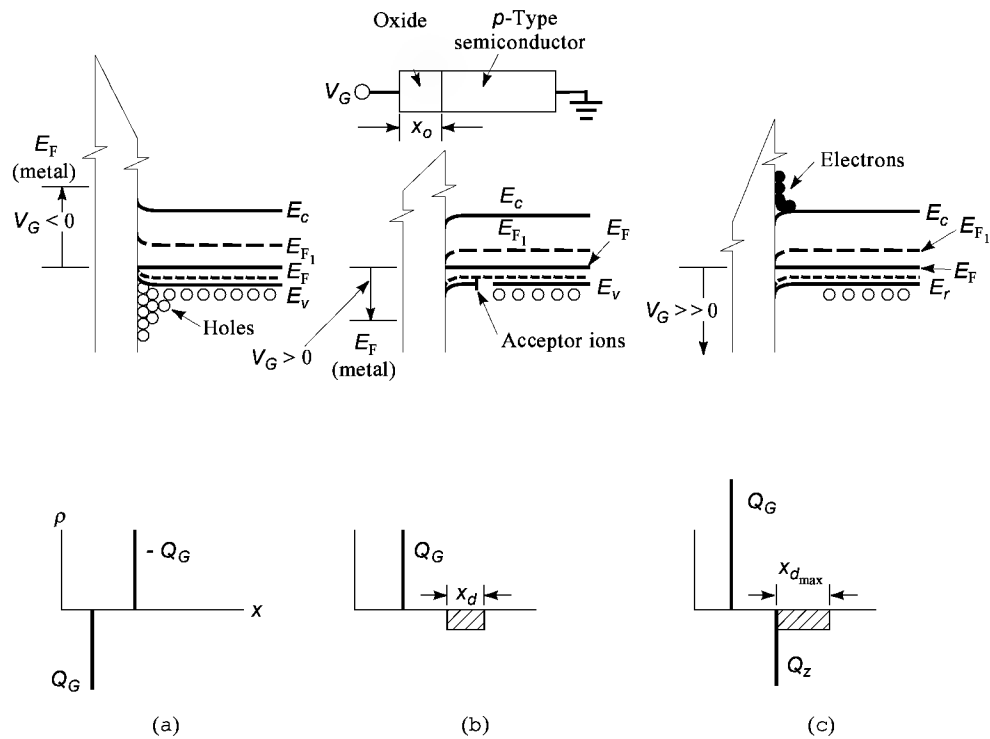


Fig. 4. Charge distributions in the voltage-biased MOS capacitor: (a) accumulation of majority carriers near surface; (b) depletion of majority carriers from surface; (c) inversion, accumulation of minority carriers near surface (26). V_G = gate voltage; Q_G = gate charge; x_d and $x_{d_{max}}$ = depletion widths; and ρ = charge density.

Thus, this MOS capacitor has three regions of operation: accumulation, for negative voltages when mobile holes are drawn to the surface; depletion, for positive voltages which repel mobile holes from the surface; and inversion, for positive voltages greater than the threshold voltage when mobile electrons are drawn to the surface. The foregoing description is of a NMOS capacitor where the semiconductor is p -type; a similar account can be given for a PMOS capacitor where the semiconductor is n -type, except that the metal plate voltages then have the reverse sign and the roles of electrons and holes are interchanged. The magnitude of the capacitance depends on voltage, as shown in Figure 5, approaching the oxide capacitance in accumulation and inversion and dropping to about 0.4 of the oxide capacitance in depletion, because the negative charge due to ionized acceptors is farther from the metal plate. The rise of capacitance in inversion assumes that electrons can follow the changes in the electric field. This only happens at low frequencies of 5–100 Hz for the M–SiO₂–Si capacitor. At high frequencies depletion continues and the capacitance continues to drop with increasing voltage reaching a minimum value corresponding to the maximum width of the depletion layer.

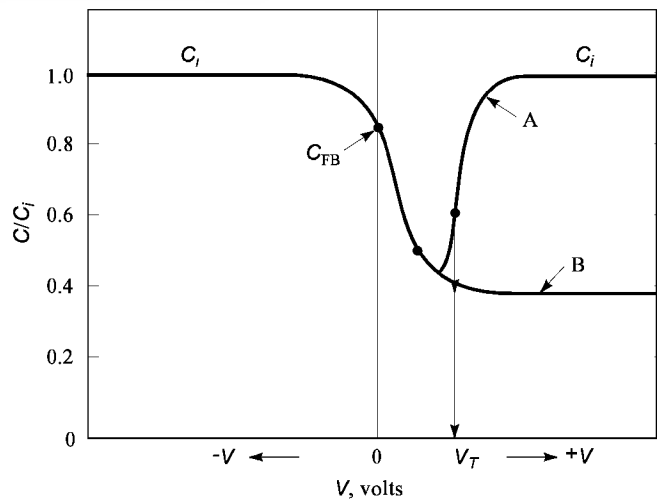


Fig. 5. NMOS capacitance voltage characteristics where C_i is the oxide capacitance, A shows low frequency characteristics, and B shows high frequency characteristics. At low frequencies C approaches C_i for negative voltages (accumulation) and positive voltages (inversion). In the flat-band (FB) condition there is no voltage difference between the semiconductor's surface and bulk. The threshold voltage, V_T , for channel formation is the point where the semiconductor's surface is as strongly n -type as the bulk is p -type (26).

The Si–SiO₂ Interface

The simple picture of the MOS capacitor presented in the last section is complicated by two factors, work function differences between the metal and semiconductor and excess charge in the oxide. The difference in work functions, the energies required to remove an electron from a metal or semiconductor, is $q\phi_{ms} \approx 25$ meV for an aluminum metal plate over a 50-nm thermally grown oxide on *n*-type silicon with $n = 10^{16}$ cm⁻³. This work function difference leads to a misalignment of energy bands in the metal and semiconductor which has to be compensated by a variation of the energy band with distance. When there is no misalignment the flat-band condition results.

Excess charge in the oxide has the effect of charge on the metal plate but has a larger effect because it is closer to the silicon surface. There are several components of excess charge. Interface trapped charges at the Si–SiO₂ interface actively exchange with carriers at the semiconductor surface. Fixed interface charges at or near the interface are immobile in an applied electric field. Oxide trapped charges which are also immobile can be created by irradiation or hot-electron injection. They have an important influence on the wear-out and reliability of MOS devices. Mobile ionic charges such as Na⁺ ions are notorious because they shift threshold voltages as they move under bias-temperature aging.

Interface states played a key role in the development of transistors. The initial experiments at Bell Laboratories were on metal–insulator–semiconductor (MIS) structures in which the intent was to modulate the conductance of a germanium layer by applying a voltage to the metal plate. However, only ~10% of the induced charges were effective in charging the conductance (3). It was proposed (2) that the ineffective induced charges were trapped in surface states. Subsequent experiments on surface states led to the discovery of the point-contact transistor in 1948 (4).

At a semiconductor–vacuum interface, dangling bonds extend from the surface atoms into the vacuum. Surface reconstruction minimizes the electronic energy at the semiconductor surface by forming bonds, but the presence of a large number of surface states can completely overshadow any effects of doping near the surface. Thus, it is important to passivate the surface by removing most of the surface states. It was noted that thermal oxidation could passivate the silicon surface and the first silicon MOSFET was created in 1960 (10). Electrons are strongly bound in the interfacial Si–O bonds and do not participate in conduction. Thermal oxidation can reduce the number of surface states by four orders of magnitude.

The excellence of a properly formed SiO₂–Si interface and the difficulty of passivating other semiconductor surfaces has been one of the most important factors in the development of the worldwide market for silicon-based semiconductors. MOSFETs are typically produced on <100> silicon surfaces. Fewer surface states appear at this Si–SiO₂ interface, which has the fewest broken bonds. A widely used model for the thermal oxidation of silicon has been developed (31). Nevertheless, despite many years of extensive research, the Si–SiO₂ interface is not yet fully understood.

Device Physics

The ideal rectifier or diode is a two-terminal device that allows current flow in only one direction. The transistor is a three-terminal device in which current flow through two terminals is controlled by the third. Transistors can be used as analogue amplifiers or digital switches.

P–N JUNCTIONS

In addition to its use as a rectifier, the *p–n* junction (26) is the fundamental building block for bipolar, junction FET (JFET), and MOSFET transistors. A thorough understanding of *p–n* junctions explains much of transistor behavior. The theory (5) of the *p–n* junction and its role in bipolar transistors was presented within a year of the discovery of the point-contact transistor.

At an abrupt interface between *n*-type and *p*-type semiconductors there is an enormous difference between electron and hole concentrations on the two sides of the interface. This corresponds to a difference in Fermi energies on the two sides. Thermal equilibrium with no applied voltage and no current flow requires a constant Fermi energy throughout the material. This equilibrium is achieved if electrons diffuse into the *p*-region and holes diffuse into the *n*-region where they become minority carriers. This leaves behind a depletion region at the interface with ionized acceptors on the *p*-side and ionized donors on the *n*-side. This charge separation leads to a built-in or diffusion voltage difference, V_{bi} , across the interface. This built-in voltage corresponds to a difference in band energies as shown in Figure 6. Because $n_{no}p_{no} = n_i^2 = n_{po}p_{po}$, where n_{no} and p_{po} are the equilibrium majority concentrations in the *n*-region and *p*-region, respectively, the equation for V_{bi} allows the determination of the concentrations

$$qV_{bi} = KT \ln(n_{no}p_{po} / n_i^2) \approx KT \ln(n_d n_a / n_i^2)$$

(8)

of minority carriers on either side of the depletion region. Thus, equation 9 can be written as follows:

$$n_{po} = n_{no} \exp(-qV_{bi} / KT)$$

(9)

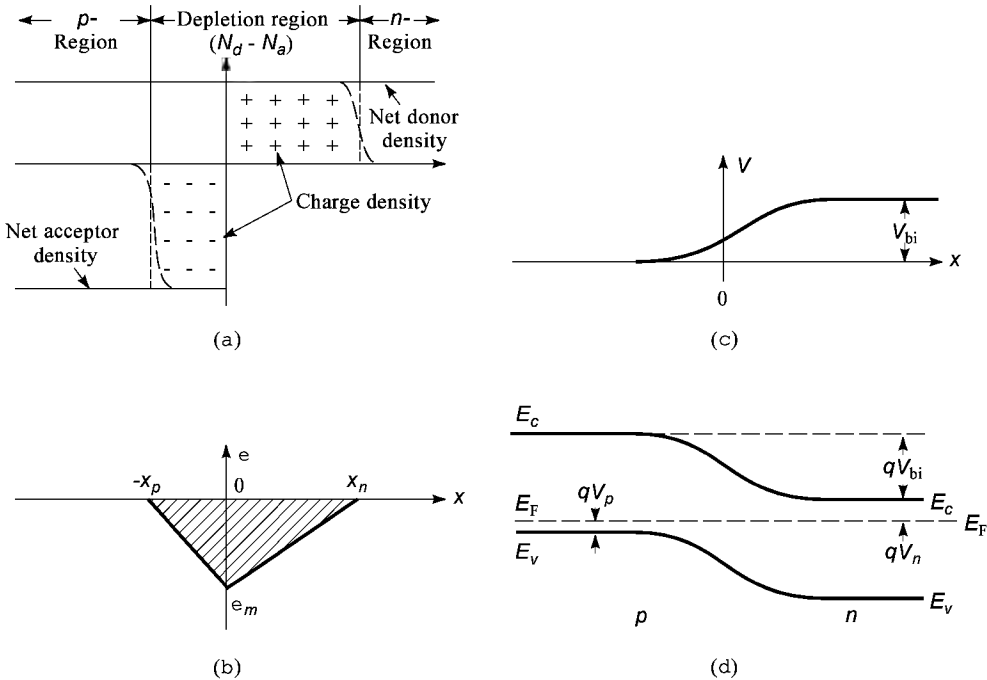


Fig. 6. An abrupt *p–n* junction in thermal equilibrium: (a) space–charge distribution where () indicate majority carrier distribution tails and the charge

density is the result of unneutralized impurity ions; **(b)**, where ϵ_m is the maximum electric field in the depletion region, and **(c)** spatial variations of the electric field and corresponding potential where V_{bi} is built-in potential or voltage and $\mathbb{A}$ is the area of diffusional potential; **(d)** energy band diagram where qV_p and qV_n are energies of ionized acceptors, a , and donors, d , relative to the band edges E_i and E_p , respectively. Where the band edges are flat, there is no electric field and the potential does not change (neutral regions) (26).

In general, in a planar process, p - n junctions are formed just below the surface of a silicon wafer by the implantation of donor ions into a p -type region or acceptor ions into an n -type region. Thus, the general concern is with $n^+ - p$ or $p^+ - n$ junctions. As the initial wafer concentration of acceptors or donors in silicon increases from 10^{14} to 10^{18} cm^{-3} , V_{bi} increases from about 0.81 to 1.04 V for a $p^+ - n$ junction and is about 10 mV higher for an $n^+ - p$ junction.

When a positive voltage difference from p to n is applied across the p - n junction, the voltage drop is mostly across this depletion region and reduces the built-in voltage. A forward voltage drop decreases the width of the depletion region as it reduces the charge imbalance required to sustain the built-in voltage. In simple models of junction behavior, voltage drops are neglected across the neutral regions, and the n -regions and p -regions outside the depletion region. Actually, there are resistive voltage drops across these regions which are proportional to current flow. This provides a parasitic resistance which needs to be included in a more accurate device model.

When electrons are injected as minority carriers into a p -type semiconductor they may diffuse, drift, or disappear. That is, their electrical behavior is determined by diffusion in concentration gradients, drift in electric fields (potential gradients), or disappearance through recombination with majority carrier holes. Thus, the transport behavior of minority carriers can be described by a continuity equation. To derive the p - n junction equation, steady-state is assumed, so that $\partial n_p / \partial t = 0$, and a neutral region outside the depletion region is assumed, so that the electric field is zero. Under these circumstances, the continuity equation reduces to a diffusion equation (eq. 10), where τ_p is the lifetime of minority

$$D_e \frac{\partial^2 n_p}{\partial x^2} - \frac{n_p - n_{po}}{\tau_p} = 0 \tag{10}$$

carriers in the p -type material. The electron current density is given by equation 11.

$$J_e = q D_e \frac{\partial n_p}{\partial x} = \frac{q D_e n_{po}(0) \left(e^{eV / K T} - 1 \right)}{L_p} \tag{11}$$

$L_p = (D_e \tau_p)^{1/2}$ is the minority carrier diffusion length for electrons in the p -region. $n_{po}(0)$ is the minority carrier concentration at the boundary between the depletion layer and the neutral region. The sign of this equation indicates that electron injection into the p -region results in a positive current flow from p to n as shown in Figure 7.

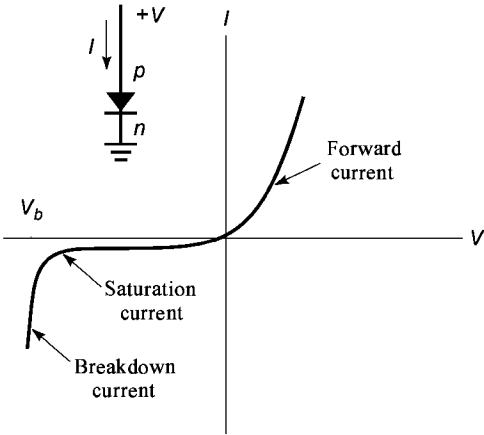


Fig. 7. Circuit symbol and current, I , voltage, V , characteristic of a p - n junction diode where V_b = breakdown voltage(32).

Hole injection into the n -region similarly results in a positive current flow from p to n^+ , leading to the Shockley diode equation (eq. 12), with a reverse

$$I = I_s \left(e^{qV / K T} - 1 \right) \tag{12}$$

saturation current, I_s , where A is the area of the junction.

$$I_s = q A \left(\frac{D_e n_{po}}{L_p} + \frac{D_h p_{no}}{L_n} \right) \tag{13}$$

For positive voltages the current, I , can become exponentially large and is limited by junction heating and burnout. A forward-biased silicon p - n junction has the voltage drop shown in equation 14. For negative voltages, $I \approx -I_s$. For

$$V = (q / K T) \ln(I / I_s) \approx 0.7 \text{ V} \tag{14}$$

an $n^+ - p$ junction, the lever rule gives $p_{no} \ll n_{po}$ so that $I_s \approx q A (D_e n_{po} / L_p)$. For reasonable values, $\mu \approx 1000 \text{ cm}^2 / (\text{V} \cdot \text{s})$, $n_a = 10^{16} \text{ cm}^{-3}$, $\tau_p = 1 \text{ }\mu\text{s}$, $L_p \approx 50 \text{ }\mu\text{m}$ and the reverse saturation current would be $17 \times 10^{-12} = 17 \text{ pA}$ for a square centimeter of junction area. Typical reverse saturation currents are about one thousand times greater as a result of generation-recombination currents in the depletion region (9). As the reverse voltage bias increases, the field increases in the depletion region until avalanche breakdown occurs, resulting in the characteristic shown in Figure 7.

The depletion region width and minority carrier distributions near the junction must be altered in order to change the voltage across a p - n diode. This corresponds to two capacitive components, the junction capacitance which dominates in reverse bias and the diffusion or charge-storage capacitance which dominates in forward bias. The charge separation caused by the concentrations of ionized donors and acceptors in the depletion region forms a capacitor. Applied voltage changes, dV , across the depletion region change the width of the depletion region. This changes the charge it contains, dQ , which then defines the junction capacitance, $C = dQ / dV$. Because the depletion layer width depends nonlinearly on the applied voltage, this capacitance

also is voltage dependent. In addition to the junction capacitance, the injection and removal of minority carriers from the neutral regions has electrical effects similar to capacitance and defines the diffusion capacitance. This capacitance depends on the lifetime of the minority carriers. Doping the $p-n$ junction with gold has been used to reduce the minority carrier lifetime, thereby reducing the diffusion capacitance and decreasing the switching time associated with the $p-n$ junction.

Avalanche or tunnel breakdown occurs in the $p-n$ junction for the large electric fields in the depletion region for sufficiently large reverse bias. Increasing the doping on either side of the junction decreases the depletion width and lowers the avalanche breakdown voltage, although the actual breakdown field increases. As the electric field increases the slope of band energy with distance increases. When the distance between conduction and valence bands at constant band energy is <10 nm, electrons can tunnel from the valence to the conduction band. As shown in Figure 8, the energy barrier through which the electrons tunnel is essentially triangular, corresponding to Fowler-Nordheim tunneling. For doping levels $>10^{18}$ cm⁻³ in silicon the critical field for avalanche breakdown exceeds the threshold for tunneling and tunnel breakdown occurs. Junction avalanche breakdown can be useful because it can be nondestructive. The junction breakdown voltage can be used as a reference voltage as in Zener diodes. Avalanche breakdown is also used for programming nonvolatile memories by hot-carrier injection. On the other hand, hot-carrier injection is a wear-out mechanism for MOSFETs.

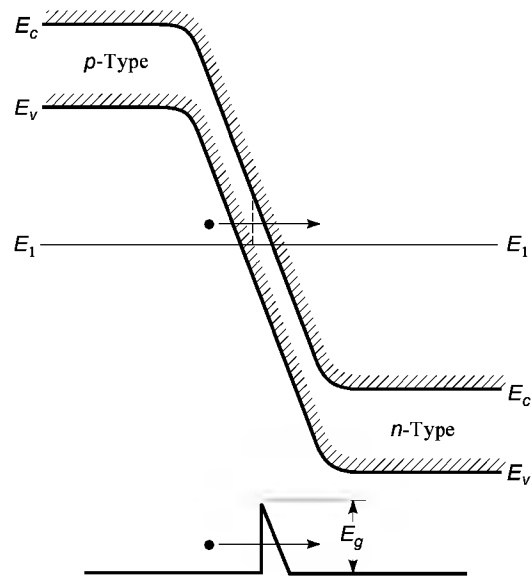


Fig. 8. Band-to-band tunneling in a semiconductor where the triangular barrier leads to Fowler-Nordheim tunneling (33).

Schottky Diodes and Ohmic Contacts. A metal–semiconductor junction may be rectifying or ohmic. Such a rectifying junction (Schottky diode) has a nonlinear, asymmetric dependence of current on voltage like the $p-n$ junction, which could be called a Shockley diode (26). An ohmic contact has a linear, symmetric dependence of current on voltage. Figure 9 shows band diagrams for Schottky barrier junctions. In equilibrium, metal and semiconductor Fermi levels must align, resulting in a Schottky barrier height $q\phi_{BO} = q\phi_M - q\chi$ at the interfaces. $q\phi_M$ and $q\chi$ are the metal work function and electron affinity, the energies required to remove an electron from the Fermi energy in the metal and the conduction band edge in the semiconductor, respectively.

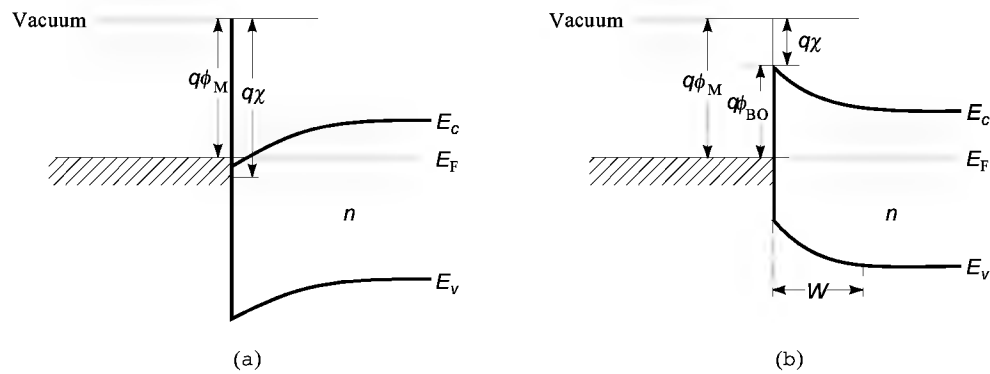


Fig. 9. Schottky barrier band diagrams: (a) a rare situation where the metal work function $q\phi_M$ is less than the semiconductor electron work affinity $q\chi$, resulting in an ohmic contact; (b) normal Schottky barrier with barrier height $q\phi_{BO}$. When the depletion width W is <10 nm, an ohmic contact forms.

Both ohmic and rectifying behavior are possible, depending on the sign of $q\phi_{BO}$. Unlike the $p-n$ junction the current in a rectifying Schottky barrier largely is controlled by thermal emission of majority carriers over the barrier. When there is no voltage difference between the metal and the semiconductor there is no current flow, ie, the electron flow from metal to semiconductor balances that from semiconductor to metal. A positive voltage applied to the semiconductor raises the barrier to electron flow from semiconductor to metal but does not affect that from metal to semiconductor. This leads to the same form of current voltage characteristics as the $p-n$ junction (eq. 12), but there are differences. The Schottky diode has a larger reverse saturation current and its switching speed is not limited by the minority carrier lifetime. Consequently, there is a large difference between forward and reverse currents (rectifying behavior) for a positive barrier (Fig. 9b), but very little difference (ohmic behavior) for a negative barrier (Fig. 9a) which does not constrain electron flow.

Because rectifying barriers are more common than ohmic barriers, dopant concentrations of about 10^{19} cm⁻³ are typically used to form ohmic contacts on silicon. The distance over which band bending occurs decreases as the semiconductor doping increases because more dopants allow more rapid depletion. Tunneling through the depletion region is possible if it is $< \sim 10$ nm, leading to ohmic behavior with a rectifying barrier.

BIPOLAR TRANSISTORS

As Figure 10 shows, the $n-p-n$ bipolar junction transistor (BJT) may be regarded as two back-to-back $p-n$ junctions separated by a thin base region (26,32,33). If external voltages are applied so that the base-emitter (BE) junction is forward biased and the base-collector (BC) junction is reverse biased, electrons injected into the base from the emitter can travel to the base-collector junction within their lifetime. If the time for minority carrier electrons to

transit the base (base transit time) is short compared with their lifetime, α , the fraction of electrons which reaches the collector junction is nearly one. When they reach the collector they are accelerated across the depletion region, giving rise to a current, $I_C = \alpha I_E \approx I_E$. Kirchoff's current law requires that the current flowing through the emitter contact must be supplied by currents flowing through the base and collector contacts, $I_E = I_C + I_B$. Substituting this equation for I_E gives a formula for I_C in terms of I_B , $I_C = \alpha I_B / (1 - \alpha) \equiv \beta I_B$. If $\alpha = 0.99$, then $\beta = 100$, and there is a substantial current gain from the base to the collector. This simple picture explains most BJT behavior. So long as the BE and BC junctions are forward and reverse biased, respectively, the bipolar transistor is in the active region with a current gain, β . The base-emitter junction must be forward biased by 0.5–0.6 V for appreciable current to flow and the BE input characteristics are those of a $p-n$ junction. Thus, emitter current depends exponentially on V_{BE} according to the diode equation (eq. 12). This diode law holds over a very wide range of currents from nanoamperes to milliamperes. Because β may vary by a factor of three over this range, the BJT may be best understood as a transconductance amplifier, where I_C is determined by V_{BE} , rather than a current amplifier, where I_C is determined by I_B . This transconductance amplifier perspective is particularly helpful when considering the design of differential amplifiers (34).

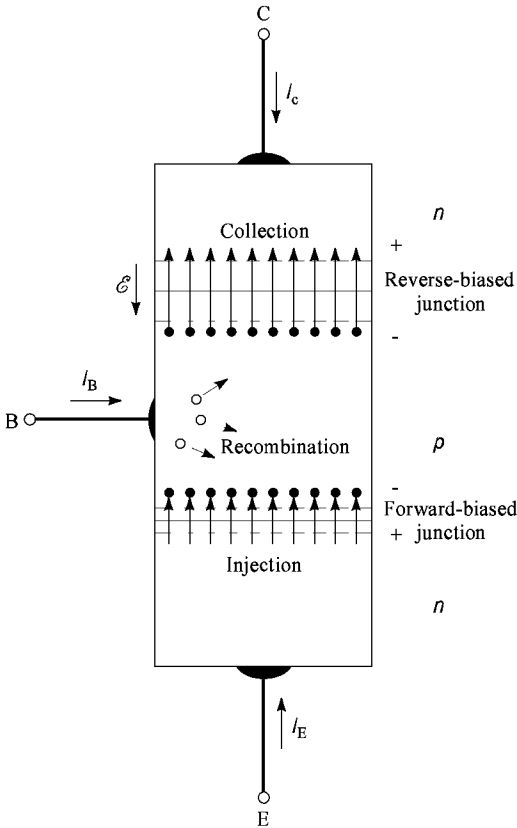


Fig. 10. The $n-p-n$ transistor biased in its active region, where I = current, () indicate depletion regions at the $p-n$ junctions, and $\mathcal{E}$ is the electric field; their width is reduced for the forward biased base-emitter (BE) junction and increased for the reverse biased base-collector (BC) junction. Note the directions of electron (•) and hole (°) flow (33).

Figure 11 shows a schematic and collector characteristics for a common emitter $n-p-n$ transistor circuit. The load line crossing these characteristics shows the allowed operation of the transistor with a supply voltage, $V_{CC} = 12\text{ V}$; a load resistor, $R_C = 2\text{ k}\Omega$, and a bias resistor, $R_B = 20\text{ k}\Omega$. The load line corresponds to the equation $V_{CC} = I_C R_C + V_{CE}$. Plotting the load line on the collector characteristics defines BJT behavior; $V_B > V_{BE} \approx 0.6\text{ V}$ is required for the BE junction to be forward biased and $I_B > 0$. For lower values of V_B , $I_B \approx 0$ and the BJT is in the cutoff region with $V_{CE} = 12\text{ V}$. As I_B increases, the transistor is in the active region until it reaches the saturation region at around 6 mA. At this point increasing the base current causes no increase in collector current (CC) so $I_B < I_C / \beta$ and V_{CE} is low; in saturation both the BC and CE junctions are forward biased.

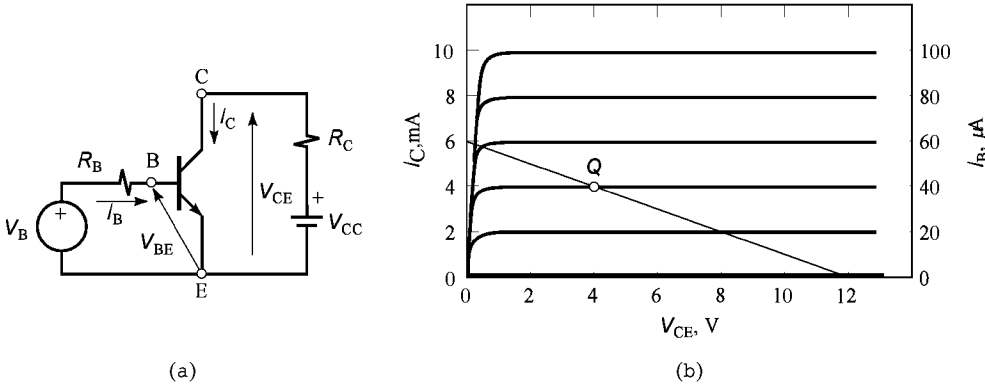


Fig. 11. (a) Common emitter circuit and (b) output characteristics for the $n-p-n$ transistor, where Q is the operating (quiescent) point as determined by R_B . See text (35).

Courtesy of Addison-Wesley Publishing Co., Inc.

A high gain transistor requires α nearly equal to 1. In the absence of collector junction breakdown, α is the product of the base transport factor and emitter efficiency. The base transport factor, α_T , is the fraction of the minority current (electrons for an $n-p-n$ transistor) that reaches the collector. $\alpha_T \approx 1 - W^2 / 2L_B^2$, where W is the base width, is the distance between emitter and collector junctions and L_B is the minority carrier diffusion length in the base. High gain transistors require a thin base as well as a long minority carrier lifetime for a large L_B . Because α_T is >0.995 in modern transistors, there is little room for improvement. The emitter efficiency, the fraction of emitter current due to minority carriers injected into the base instead of the emitter,

must be maximized for high gain transistors. Under these conditions, β is proportional to $n_E p_B W$ where n_E and p_B are the emitter and base dopings. To design a transistor for maximum β the Gummel number, $p_B W$, must be minimized (33).

The frequency response or switching speed of the bipolar transistor is governed by the same processes which control the speed of the $p-n$ junction, the capacitance associated with the movement of charge into and out of the depletion regions. To achieve high frequencies the dimensions of the active areas and parasitic circuit elements must be reduced. The two critical dimensions are the width of the emitter contact and the base thickness, W . The cutoff frequency, f_T , is the frequency at which $h_{ie} \partial I_C / \partial I_B = 1$; $f_T = (2\pi\tau_{ec})^{-1}$, where τ_{ec} is the emitter-to-collector delay time and is the sum of the emitter depletion layer charging time, the base layer charging time, and the collector charging time (26). To increase f_T the transistor should have a thin base, a narrow collector width, and be operated at high I_C . However, decreasing collector width decreases the breakdown voltage for the BC junction. The minimum base width (thickness) is determined by the depletion region widths of the BE and BC junctions. Narrow base widths or low base dopings allow the collector and emitter depletion regions to come in contact above some value of V_{BC} . This condition is called punchthrough and leads to large currents uncontrolled by the base.

To achieve the lowest possible delay a bipolar switching transistor developed by IBM minimizes parasitic resistances and capacitances. It consists of self-aligned emitter and base contacts, a thin intrinsic base with an optimized collector doping profile, and deep-trench isolation (36). Devices must be isolated from each other to prevent unwanted interactions in integrated circuits. While $p-n$ junctions can be used for isolation, IBM's approach etches deep trenches in the silicon wafer which are filled with SiO_2 to provide electrical insulation.

FIELD-EFFECT TRANSISTORS

Unlike bipolar transistors, field-effect transistors (FETs) are unipolar devices whose characteristics are determined by majority carrier transport through a channel (26,33,37). Current flow in this channel is altered by forming or constricting current flow by the application of an electric field. In the junction FET (JFET), a conducting channel between two ohmic contacts is constricted by the depletion regions of two opposing $p-n$ junctions. In the MOSFET, a conducting channel between the two $p-n$ junctions is formed by application of a field at the silicon surface by a MOS capacitor. MOSFET operation is considered here at length because it dominates silicon device technology.

A cross section of a MOSFET is shown in Figure 12. The MOSFET is formed by growing a thin thermal oxide on the surface of a silicon wafer and depositing a polysilicon gate which serves as the MOSFET's metal plate. This is a self-aligned structure in which the polysilicon gate acts as a mask during subsequent wafer processing to define the implantation source and drain regions. For an NMOSFET the silicon substrate is p -type and the source and drain regions are n^+ , forming back-to-back diodes. When a gate voltage, V_{GS} , is applied which is positive in relation to the source, the surface is influenced as it would be in an MOS capacitor. If V_{GS} is larger than a threshold voltage, V_T , a surface channel forms by inversion of the surface. An important difference from the usual MOS capacitor is that the n^+ source and drain junctions are a ready supply of electrons for the channel. Thus, the channel can be formed at high speed even at cryogenic temperatures because the source and drain regions are degenerately doped.

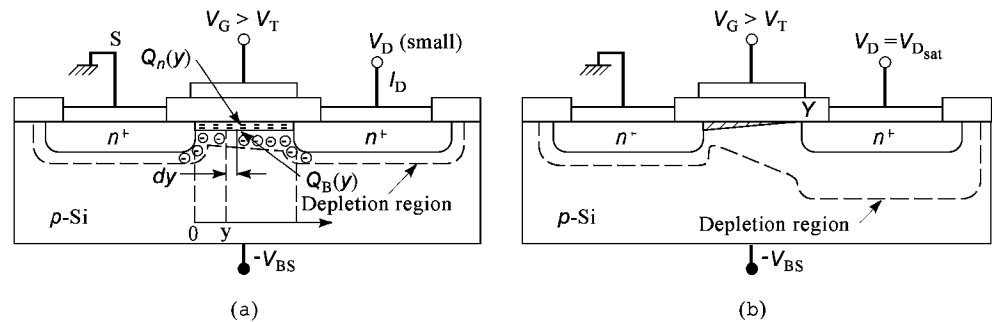


Fig. 12. NMOSFET operated (a) in the linear (low drain voltage) region, where S is the grounded source contact, y is the distance from source to drain, and d is differential increase and (b) at the onset of the saturation region, $V_D = V_{Dsat}$. The point Y indicates the channel pinch-off point. For $V_D > V_{Dsat}$ point T moves to the left, reducing the effective channel length, but the drain current remains nearly constant. See text (26).

The MOSFET has three regions of operation. The cutoff region occurs for $V_{GS} < V_T$. In this region, the drain-to-source current is the reverse saturation current of the back-to-back source and drain junctions. This leakage current is small but nonzero and allows charge to leak off capacitors which are isolated by cutoff MOSFETs. Because this is how bits are stored in dynamic memory (DRAM) cells, DRAMs must be regularly refreshed to retain their memory.

When $V_{GS} > V_T$ the MOSFET conducts. The conduction current is determined by Q/t_t where Q is the amount of charge in the inversion layer and t_t is the transit time for electrons to travel from source to drain. $Q = C_o' LW(V_{GS} - V_T)$ where $C_o' = \epsilon_{ox} \epsilon_o / T_{ox}$ is the gate oxide capacitance per unit area and L and W are the length and width of the channel, respectively. $t_t = L / v_D = L^2 / \mu V_{DS}$ where v_D is the electron's drift speed, μ is the electron's mobility in the channel, and V_{DS} is the drain-to-source voltage. A more accurate expression for conduction in the resistive or linear region for low V_{DS} , which takes into account the reduction in gate voltage near the drain, is equation 15:

$$I_{DS} = \mu C_o' \left(\frac{W}{L} \right) \left[(V_{GS} - V_T) - \frac{V_{DS}}{2} \right] V_{DS} \tag{15}$$

The channel conductance ($\approx I_{DS} / V_{DS}$) is proportional to the ratio of width to length for the channel, $W:L$. $\mu C_o'$ is determined by the fabrication process, and L by photolithography leaving the layout ($W:L$) to be adjusted by the circuit designer to achieve an appropriate trade-off between circuit speed and area. For V_{DS} small compared with $V_{GS} - V_T$, the channel conductance is a constant, increasing with $V_{GS} - V_T$.

As V_{DS} increases, the depletion width at the drain junction grows and can accommodate more charge. Thus, less charge is needed in the inversion layer to balance the gate charge. Because the surface potential at the drain edge of the channel is $V_{D\phi}$, when $V_{GS} - V_{DS} < V_T$ inversion can no longer be maintained and the inversion layer is pinched-off at the drain end. For $V_{DS} > V_{GS} - V_T$, the MOSFET is in the saturation region and the saturation current I_{DS} is given by equation 16. For $V_{DS} > V_{GS} - V_T$ the pinched-off region can accommodate

$$I_{DS} = \frac{\mu C_o'}{2} \left(\frac{W}{L} \right) (V_{GS} - V_T)^2 \tag{16}$$

a large potential drop, which fixes I_{DS} at its saturated value. If the electric field across the channel, V_{DS} / L , is large enough the drift velocity may reach its saturation value, v_s , and equation 16 becomes equation 17.

$$I_{DS} = \left(\frac{WC'_o v_s}{2} \right) (V_{GS} - V_T)$$

(17)

Several features should be noted about MOSFET operation. One is that MOSFETs are horizontal devices whose currents are limited by the depth of the channel. As a result, MOSFETs do not have as much current drive as bipolar transistors which, as vertical devices, have more area through which to pass current. This is less true for JFETs where current flow is not restricted to a thin surface channel. The discussion so far has tacitly assumed that $V_T > 0$, corresponding to an enhancement mode NMOSFET or NFET. Suitable channel doping can cause $V_T < 0$, corresponding to a depletion mode NMOSFET. Depletion mode MOSFETs conduct when $V_{GS} = 0$ and are used to provide pull-up resistors in NMOS technology. The MOSFET is a symmetrical device which can conduct equally well in either direction. This allows it to be used in switch or pass-transistor logic as well as static and dynamic logic families.

The previous discussion about NMOSFETs applies equally well to PMOSFETs or PFETs. PMOSFET channels have hole conduction between p^+ drain and source regions implanted into an n -type substrate. $V_T < 0$ and $V_{GS} < V_T$ for conduction. Because the hole mobility is approximately half the electron mobility, PFETs provide about half the gain for a given device width. Designers can compensate for this in the layout of a circuit by making PFET gate widths twice as wide as NFET gate widths.

NFETs and PFETs can be combined to form complementary MOS (CMOS) circuits (38). Figure 13 shows the circuit diagram for a CMOS inverter, the basic building block for CMOS logic, and the corresponding device cross section. It is customary to choose $V_{SS} = 0$ (ground) as the NFET source and V_{DD} (supply line) as the PFET source. Until recently, corresponding to gate lengths $> 1 \mu\text{m}$, the supply voltage has been $V_{DD} = 5 \text{ V}$. The cross section, which is roughly to scale, is for a p -well process in which the NFET's n^+ source and drain regions are implanted into a p -type well formed by implantation of acceptors into an n -type substrate. The scale of a process is determined by the length of the polysilicon gates which is $2 \mu\text{m}$ for a $2\text{-}\mu\text{m}$ process. Note that the thickness of the gate oxide (40 nm for a $2\text{-}\mu\text{m}$ process) is much less than the other thicknesses of material.

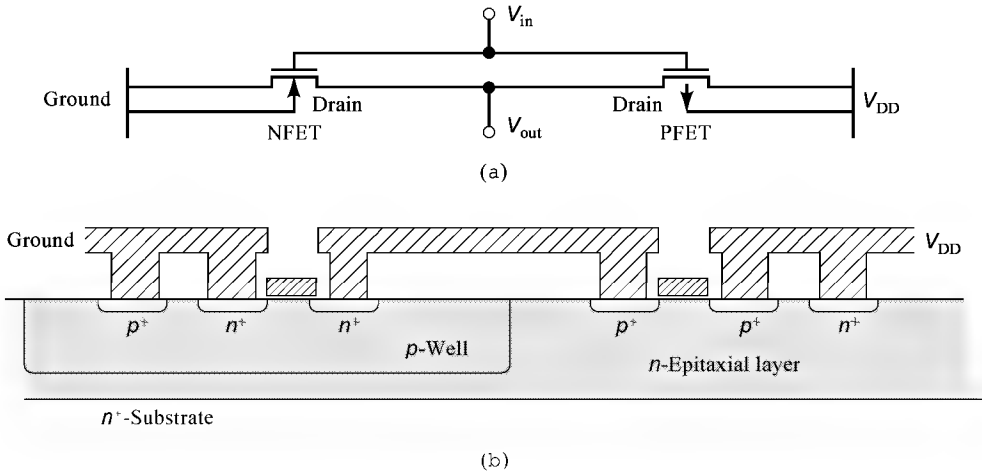


Fig. 13. (a) The CMOS inverter circuit. The FET circuit symbols emphasize that MOSFETs are actually four-terminal devices in which the n -substrate is connected to V_{DD} for the PFET and the p -substrate is connected to the ground for the NFET. Note the conventions on drain location for the PFET and NFET. (b) Corresponding cross-sectional view roughly to scale for a $2\text{-}\mu\text{m}$ CMOS process, where represents silicon, SiO_2 , polysilicon, and aluminum.

When 5 V is applied to the input, the NFET is turned on and the PFET is cut off. The pull-down current flowing through the NFET pulls down the output to zero. Because the output is connected to the gates of subsequent MOSFETs, the output sees a capacitor. Thus, pulling down the output to zero requires discharging this capacitor. If the output was originally 5 V, the NFET is in the saturation region and maximum current flows. As the output drops, the NFET enters the resistive region, less current flows, and this increases the fall time of the output. Similarly, when 0 V is applied to the input, the NFET is cut off and, because $V_{GS} = 5 \text{ V}$, the PFET turns on and the pull-up current flowing through the PFET charges the capacitor load to 5 V in some rise time.

Inherent in the CMOS structure is a four-layer $p^+ n p n^+$ structure between the p^+ PFET and n^+ NFET drain diffusions. Such a structure behaves like a thyristor and can be triggered into a high current mode, called latchup, which can destroy CMOS circuits (38). Latchup can be prevented if the gains of the parasitic $p^+ n/p$ and $n/p p^+$ BJTs in the $p^+ n p n^+$ structure can be kept sufficiently low. Currents flowing through the substrate forward-bias the BE junction of the parasitic $p^+ n/p$ BJT. This voltage can be reduced by using a highly doped n^+ substrate to reduce the substrate resistance as shown in Figure 13. Reasonable MOSFETs are produced by implantation into a lower resistance n -type epitaxial layer grown on this substrate. Likewise, retrograde doping of the p -well, in which the doping increases with depth, reduces the well resistance and consequent forward biasing of the parasitic $n/p p^+$ BJT. Trench isolation, in which a deep trench is etched at the n - p boundary and filled with SiO_2 , is very effective in reducing latchup (39). This is an effective approach for companies that have already developed the technology for trench capacitors to store charge in DRAMs. In addition to using latchup-resistant CMOS processes, there are many techniques for designing CMOS circuit layouts to make them latchup resistant.

Current only flows in the CMOS inverter when the load capacitor is being charged or discharged. No current flows to maintain a logic level ($0 \sim 0 \text{ V}$, $1 \sim 5 \text{ V}$). Because power is dissipated only when current flows, the amount of power dissipation in a CMOS circuit is proportional to circuit activity, the charging or discharging of capacitors. In equation 18, N_{sw} is the number of circuits switching with load capacity C , and f is the frequency of operation (40); f would be the clock frequency for synchronous digital circuits.

$$P = N_{sw} C f V_{DD}^2$$

(18)

Bipolar circuits tend to dissipate much more power because large currents are always flowing. NMOS circuits dissipate more power than CMOS circuits because the PFET is replaced by a depletion mode NFET which always conducts. Thus, traditionally, CMOS has been regarded as a low power circuit technology. However, this is only true for circuits operating at low frequency, with low activity, or at low supply voltages. Modern microprocessor chips contain millions of transistors, operate at high frequencies, and hence dissipate substantial power. For example, Digital Equipment Corporation's (DEC) 21064 Alpha chip, introduced in 1992, contained 1.68 million transistors with $0.75 \mu\text{m}$ gate lengths and dissipated 30 W when operating at 200 MHz from a 3.3-V supply (41). The Alpha chip is mentioned because its design emphasized achieving maximum clock frequency for a given level of technology, as defined by gate length. Clearly, because power dissipation depends quadratically on supply voltage, lowering the supply voltage is the most effective way to lower power dissipation. If the DEC Alpha chip were operating at 5 V it would dissipate almost 70 W.

The lower power dissipation associated with CMOS technology makes it attractive to crowd as many FETs as possible on a chip. The remarkable increase in the number of FETs per chip has been the result of shrinking FET sizes. If the gate length, L , and width, W , are decreased by a factor of two,

four times as many FETs can be placed in the same area. Generalized scaling (42) has allowed the behavior of FETs to remain the same when their size shrinks. In constant-field scaling the electric fields are kept the same as device dimensions shrink. This allows the circuit characteristics to keep the same shape, which means that the gate oxide thickness and the depth of the source and drain junctions must shrink by the same factor, $1/\alpha$, as L and W . It also means that V_{DD} and V_T need to shrink by the same factor, $1/\alpha$. Shrinking V_T requires increasing the substrate doping by α to maintain the relative size of the depletion region. Constant field scaling reduces both load capacitance and gate delay by $1/\alpha$. Constant field scaling works with little modification down to gate lengths of $\sim 0.1\ \mu\text{m}$.

The desire to maintain 5 V system voltages has led to a modified, constant voltage scaling in which the supply voltage remained constant and gate delay was reduced by $1/\alpha^2$. However, the need to reduce power dissipation while increasing clock frequency to achieve higher performance is leading to rapid decreases in supply voltage to 3.3 V in present 0.35- μm technology, to 2.5 V with 0.25- μm technology, and perhaps to 1.8 V with 0.18- μm technology. This is particularly true for applications such as laptop computers where low power simplifies packaging requirements and increases battery life.

As device dimensions shrink, subthreshold leakage becomes an increasing problem. I_{DS} does not cut off sharply at V_T . Rather, around V_T , I_{DS} depends exponentially on V_{GS} , requiring much lower V_{GS} to reduce I_{DS} to the point where a high enough output voltage is assured. Because subthreshold behavior depends on the carrier energy distribution, it is substantially reduced at cryogenic temperatures, leading to much sharper FET turnoffs.

BiCMOS processes combine bipolar and MOSFET devices in a more complex fabrication process to provide the benefits of both technologies. For digital circuits a buried layer and epitaxial layer are combined with the wells of a CMOS process to allow the fabrication of $n\text{-}p\text{-}n$ transistors. In a BiCMOS chip, CMOS provides a high logic density while the $n\text{-}p\text{-}n$ transistors at the output of gates provide a higher drive for output pins and other high capacitance loads such as buses and clock trees. Mixed-signal ICs combine analog and digital functions. They require a more complex BiCMOS process to fabricate both $n\text{-}p\text{-}n$ and $p\text{-}n\text{-}p$ transistors for operational amplifiers (op amps) and other linear (small signal) circuits.

Newer Applications

Although the predominant applications for silicon-based semiconductors are transistors and integrated circuits (ICs) for computation and communication, several other applications are of growing significance. These include flat-panel displays, flash memories, power semiconductors, micromechanics, and cryoelectronics. New materials technology such as SiGe or SiC may become important for higher performance or higher temperature operation (see SEMICONDUCTORS, COMPOUND). It should also be noted that silicon devices correspond to front-end processing for fabricating integrated circuits. As the minimum feature sizes of devices shrink to submicrometer dimensions and chips increase in size, back-end processing of multilayer interconnect structures is becoming much more important in determining integrated circuit performance (43). Chip speeds will be limited by interconnects even if device times shrink to zero.

Displays. Flat-panel visual displays (44,45) remain an active area for research and development. Laptop computers require an alternative to the bulky and fragile cathode ray tube (CRT) used as a video monitor. This has provided a significant market for flat-panel displays (FPDs), accounting for about 60% of a growing market of \$4.6 billion in 1993 (44). Flat-panel displays consist of an array of closely spaced, separately contacted picture elements or pixels. There are several competing technologies of which active matrix liquid crystal displays (AMCLDs) are the dominant technology for laptop computers. AMLCDs sandwich liquid crystal material between two sheets of glass. One sheet contains a matrix of thin-film transistors (TFTs) connected to transparent pixel electrodes. The other sheet contains color filters, electrodes, and a polarizer (45). The TFTs, MOSFETs made with amorphous silicon, are individually addressable and control the amount of light passing through each pixel. This precise voltage control has helped the AMCLD to be the first technology whose appearance is comparable to a CRT in resolution, contrast, color, and viewing angle, if not brightness. Many manufacturing issues for FPDs are different from those for ICs. For example, minimum feature sizes are larger ($\sim 5\ \mu\text{m}$) but must be maintained over much larger distances ($\sim 25\ \text{cm}$). This requires precise lithography control over much larger areas than single chips. Unlike memory arrays, which can include redundant cells to replace defective cells, a single defect is visible in an FPD. Present prices of 26 cm color AMLCDs are about three times the prices of 35 cm color CRTs (44), providing considerable incentive to improve manufacturing productivity.

Nonvolatile and Flash Memories. Nonvolatile memories (NVM) retain their memory state in the absence of a supply voltage, unlike static random access memories (SRAMs), which require a supply voltage to maintain their memory state, or dynamic RAMs (DRAMs) which require a supply voltage and regular refresh of the memory states (46,47). The memory states of nonvolatile, read-only memories (ROMs) are set during manufacture by the presence or absence of a MOSFET. The electrically programmable ROM (EPROM) (48) stores the memory state in a MOSFET by storing charge on a floating gate completely surrounded by SiO_2 and not electrically connected to the surrounding circuitry. Floating gate devices are nonvolatile because the floating gate can retain charge for over 10 years when stored at 125°C. Charge is stored (written or programmed) on the floating gate by avalanche injection from the drain junction. Charge is erased in the EPROM by shining uv light on the floating gate to give its electrons enough energy to escape. The memory state can be read rapidly by determining whether the floating gate MOSFET is conducting or cut off. Electrically erasable PROMs (EEPROMs) use Fowler-Nordheim tunneling between the floating gate and the substrate to erase the floating gate. A significant disadvantage of EPROMs is the 20 to 30 minutes required for erasing and reprogramming. EEPROMs are more expensive because they have more complexity and correspondingly lower density, as well as reliability problems. Flash memories are lower cost and allow quick (1 s) bulk erasure of memory blocks. They are ideal for most memory applications which require frequent updates. In 1994 the worldwide nonvolatile memory market was about \$2.7 billion divided roughly between 50% EPROMs, 35% flash, and 15% EEPROMs. In the future the market is anticipated to grow on the order of 20% yr and to be dominated by flash memories.

Power Semiconductors. Power semiconductor devices cover a diverse range of applications from 100 W at microwave frequencies (microwave ovens) to 100 MW at low frequencies (motor drives or power supplies) (49). Automotive electronics and switching power supplies require relatively low voltages ($<100\ \text{V}$) but high currents. Other applications follow a trajectory of increasing breakdown voltage and current handling capability. Over 1 kA with 1 kV blocking voltage are required for traction or power distribution applications. For operation at lower voltages ($<200\ \text{V}$) silicon power MOSFETs are the preferred device because of their low on-state resistance, high input impedance, fast switching speed, and excellent safe operating area under forward bias. However, the on-resistance and corresponding power dissipation of MOSFETs increases rapidly with increasing blocking voltage, making insulated gate bipolar transistors (IGBTs) preferred for higher voltage applications. At the highest current levels a MOS-gated power thyristor is preferred because of its low on-state voltage drop. The thyristor is a four-layer p^+n-pn^+ structure in which the lightly doped n -drift layer serves as the base of a $p\text{-}n\text{-}p$ BJT coupled with an $n\text{-}p\text{-}n$ BJT. Like latchup in a CMOS circuit, the thyristor can be triggered into high current conduction. A single power thyristor may occupy an entire 20-cm wafer.

Avalanche breakdown sets a fundamental limit on the maximum operating voltage for power devices. Because depletion layer curvature increases the crowding of electric field lines and reduces the breakdown voltage, multiple-junction field rings or other termination techniques are used to reduce the depletion layer curvature and increase the breakdown voltage. Large device areas are required to handle large current flows, and therefore a single power MOSFET may occupy an entire 1-cm² chip. More vertical channel structures, such as VMOS, DMOS, or UMOS (49), offer higher performance than the surface channels of conventional MOSFETs. The surface of the chip is covered with an array of parallel MOSFET cell structures.

Micromechanics. Micromechanics uses microelectronic fabrication techniques to fabricate microscopic micromechanical structures and machines (50). Chemical etching techniques have been developed which allow formation of three-dimensional shapes such as pits, trenches, holes, diaphragms, and cantilevers. Anisotropic etchants such as KOH are particularly useful in forming shapes because the etch rate depends dramatically on crystal plane. For silicon, the KOH etch rate ratio is about 35:1 for {100} : {111} crystal planes (51) and $>400 : 1$ for {110} : {111} crystal planes (52). The mechanical properties of silicon are also important when forming micromechanical structures. Although single-crystal silicon is brittle and fractures when its elastic limit is exceeded, it has a higher elastic limit than steel and remains strong under repeated cycles of tension and compression. Micromechanical applications include the fabrication of ink-jet nozzle arrays, gas chromatographs, microgrooved heat exchangers for cooling semiconductor chips, and cantilever accelerometers for automobile airbags. Techniques have also been developed for fabricating rotating gears and miniature motors.

Cryoelectronics. Operation of CMOS devices at lower temperatures offers several advantages and some disadvantages (53). Operation at liquid nitrogen temperatures (77 K) has been shown to double the performance of CMOS logic circuits (54). In part, this is the result of the increase in electron and hole mobilities with lower temperatures. The mobility decreases at high fields as carrier speeds approach saturation. Velocity saturation is more important for cryoelectronics because saturation velocities increase by only 25% at 77 K but saturation occurs at much lower fields. Although speedup can

be achieved with little or no process change, the magnitude of the threshold voltage tends to rise at cryogenic temperatures. However, subthreshold characteristics become significantly sharper, reducing the impact of subthreshold currents on submicrometer CMOS performance. As a consequence, the performance of new submicrometer MOSFETs are often studied at 77 K where satisfactory operation is easier to achieve. Almost no latchup is observed at low temperatures because the beta of the parasitic bipolar decreases so rapidly. On the other hand, hot-carrier injection becomes a more serious problem.

Additional benefits of cryogenic operation are possible if circuits are redesigned to take advantage of the exponential decrease in leakage current with lower temperatures. Thus, DRAM cells become pseudostatic, retaining their charge for many days. Similar exponential improvements in reliability are possible for Arrhenius-based degradation mechanisms. In addition, metal interconnect resistance should drop in proportion to temperature, reducing interconnect delays. The thermal conductivity of silicon also rises at low temperatures, aiding the removal of heat from active devices. However, the low efficiencies of practical refrigerators mean that much more heat must be rejected to room temperature than is generated at cryogenic temperatures. Low power design is much more important for cryoelectronics. If compact economical refrigerators can be developed, cryoelectronic CMOS could become an attractive alternative for much higher performance without scaling technology.

New Materials Technology. The unique chemical compatibility of SiO₂ with silicon and aluminum has been a significant factor in the dominance of silicon-based semiconductor technology for MOSFETs in particular and integrated circuits in general. Two enhancements of conventional bipolar or MOS processes have been studied: silicon on insulator (SOI) and SiGe. An alternative material of importance is SiC.

Silicon on Insulator. This technology makes MOSFETs using crystalline silicon films on insulating substrates. Initial SOI efforts were focused on silicon films which were heteroepitaxially grown on sapphire substrates. The more popular recent option has been to produce silicon films over SiO₂ on silicon wafers. For example, SIMOX wafers are prepared by deep ion implantation (qv) of oxygen atoms to produce a buried oxide layer. Because the insulating substrate isolates NFETs and PFETs, latchup and body-effect problems are removed. The enhanced radiation tolerance which results has been the chief justification for SOI technology. However, the absence of wells allows tighter transistor packing and the reduced substrate capacitance allows higher device speeds. The main challenges in scaling CMOS on SOI to submicrometer channel lengths have been short channel and floating body effects (55,56). SOI is an ideal choice for low voltage, hence low power, CMOS technology because of the near absence of body effects and junction capacitance (57).

Adding germanium to the device structure has been shown to markedly improve BJT performance (58). One approach has been a Si-SiGe emitter-base (EB) junction. The reduced band gap of the SiGe alloy provides a reduced barrier to electron injection into the base. This reduces hole injection into the emitter and thereby increases gain by increasing emitter efficiency, γ . This process is less compatible with conventional high performance silicon BJT processing than increasing the Ge content in the base from 0% at the EB junction. The Ge concentration gradient creates an electric field in the base which aids electron transport and thereby increases gain by reducing the base transit time, τ_b .

Silicon Carbide. Like SiGe, SiC is a compound semiconductor composed entirely of Group IV elements and is expected to have properties intermediate between those of silicon and diamond. Several polytypes of SiC exist but homoepitaxial growth on 6H-SiC substrates has been preferred. As Table 1 indicates, its high band gap leads to spectacular decreases in intrinsic carrier concentration which means that SiC devices should have excellent performance at high temperatures (27). Early SiC devices were blue light-emitting diode (LED) and uv photodiode detectors. However, the indirect band gap of SiC may limit its use for optoelectronic applications. In this and other applications, Groups 13 and 14 (III-V) compounds such as GaN are strong competitors. SiC has room temperature (300 K) mobilities of 400 and 75 cm²/Vs for electrons and holes, respectively (27). By careful optimization of the oxidation process the interface state density can be reduced below 10¹¹ eV⁻¹cm⁻² at the SiO₂-SiC interface. This allows MOSFETs to be fabricated. Although SiC MOSFETs with very low on-resistances for voltages <5000 V are possible, SiC's higher threshold voltage and lower mobility favors Si MOSFETs for voltages <200 V.

BIBLIOGRAPHY

"Semiconductors (Theory)" in *ECT* 2nd ed., Vol. 17, pp. 834-861, by P. J. Dean, Bell Telephone Labs, Inc.; "Semiconductors (Theory and Application)" in *ECT* 3rd ed., Vol. 20, pp. 601-633, by S. A. Schwartz, Bell Laboratories, Inc.

1. A. H. Wilson, *Proc. R. Soc. (London)* **A133**, 458 (1931).
2. J. Bardeen, *Phys. Rev.* **71**, 717 (1947).
3. W. Shockley and G. L. Pearson, *Phys. Rev.* **74**, 232 (1948).
4. J. Bardeen and W. H. Brattain, *Phys. Rev.* **75**, 1208 (1949).
5. W. Shockley, *Bell Syst. Tech. J.* **28**, 435 (1949).
6. W. Shockley, *Electrons and Holes in Semiconductors with Applications to Transistor Electronics*, D. Van Nostrand Co., Inc., Princeton, N.J., 1950.
7. G. L. Pearson and W. H. Brattain, *Proc. IRE*, 1794 (1955).
8. D. M. Chapin, C. S. Fuller, and G. L. Pearson, *J. Appl. Phys.* **25**, 676 (1954).
9. C. T. Sah, R. N. Noyce, and W. Shockley, *Proc. IRE* **45**, 1228 (1957).
10. D. Kahng and M. M. Atalla, *IRE Solid-State Device Resource Conference*, Philadelphia, Pa., 1960.
11. D. Kahng, *IEEE Trans. Electron Devices* **ED-23**, 655 (1976).
12. A. L. Robinson, *Science* **195**, 1179 (1977), special issue devoted to the electronics revolution.
13. P. Singer, *Semiconductor Int.*, 46 (Jan. 1995).
14. M. Fleming, *Semiconductor Int.*, 39 (Mar. 1995).
15. *Chem. Week*, 15 (Nov. 21, 1979).
16. K. S. Henry, *Semiconductor Int.*, 97 (July 1995).
17. *Semiconductor Int.*, 83 (Mar. 1996).
18. G. Moore, *IEEE Spectrum*, 30 (Apr. 1979).
19. K. Rose, *IEEE Circuits Devices*, 26 (Nov. 1991).
20. J. Armstrong, *IEEE Spectrum*, 53 (Jan. 1994).
21. H. R. Huff and F. Shimura, *Solid State Technol.*, 103 (Mar. 1985).
22. E. Yablonovitch, *Science* **246**, 347 (1989).
23. T. M. Brown, *Semiconductor Int.*, 223 (May 1987).
24. M. K. Balazs, *Solid State Technol.*, 75 (Oct. 1993).
25. S. K. Ghandhi, *VLSI Fabrication Principles*, 2nd ed., John Wiley & Sons, Inc., New York, 1994.
26. S. M. Sze, *Physics of Semiconductor Devices*, 2nd ed., John Wiley & Sons, Inc., New York, 1981.
27. M. Ruff and co-workers, *IEEE Trans. Electron Dev.* **41**, 1040 (1994).
28. C. Kittel, *Introduction to Solid-State Physics*, 6th ed., John Wiley & Sons, Inc., New York, 1989.
29. H. V. Malmstadt, C. G. Enke, and S. R. Crouch, *Electronic Analog Measurements and Transducers*, W. A. Benjamin, Inc., Menlo Park, Calif., 1973.
30. R. F. Pierret, *Advanced Semiconductor Fundamentals*, Addison-Wesley Publishing Co., Reading, Mass., 1987.
31. B. E. Deal and A. S. Grove, *J. Appl. Phys.* **36**, 3770 (1965).
32. B. G. Streetman, *Solid State Electronic Devices*, 4th ed., Prentice-Hall, Inc., Englewood Cliffs, N.J., 1988.
33. S. K. Ghandhi, *The Theory and Practice of Microelectronics*, John Wiley & Sons, Inc., New York, 1968.
34. P. Horowitz and W. Hill, *The Art of Electronics*, 2nd ed., Cambridge University Press, Cambridge, U.K., 1989.
35. A. B. Carlson and D. G. Gisser, *Electrical Engineering*, Addison-Wesley Publishing Co., Reading, Mass., 1981.
36. K. H. Brown and co-workers, *IBM J. Res. Devel.* **36**, 821 (1992).

37. A. S. Grove, *Physics and Technology of Semiconductor Devices*, John Wiley & Sons, Inc., New York, 1967.

38. N. H. E. Weste and K. Eshraghian, *Principles of CMOS VLSI Design*, Addison-Wesley Publishing Co., Reading, Mass., 1993.

39. C. W. Koburger and co-workers, *IBM J. Res. Devel.* **39**, 215 (1995).

40. K. Bernstein and co-workers, *IBM J. Res. Devel.* **39**, 33 (1995).

41. D. W. Dobberpuhl and co-workers, *Digital Tech. J.* **4**, 35 (1992).

42. R. H. Dennard and co-workers, *IEEE J. Solid State Circ.* **SC-9**, 256 (1974).

43. R. F. Sechler, *IBM J. Res. Devel.* **39**, 23 (1995).

44. P. Singer, *Semiconductor Int.*, 78 (July 1994).

45. C. Chinnock, *Solid State Technol.*, 75 (Apr. 1978).

46. S. Lai, *Proceedings of the International Symposium on VLSI Technology, Systems, and Applications (VLSITS/A)*, Taipai, Taiwan, 1993, p. 47.

47. B. Prince, *Semiconductor Memories*, Second ed., John Wiley & Sons, Inc., New York, 1991.

48. D. Frohman-Bentchokowsy, *IEEE J. Solid State Circ.* **SC-6**, 301 (1971).

49. B. J. Baliga, *Physics of Power Semiconductor Devices*, PWS Publishing Co., Boston, Mass., 1995.

50. J. B. Angell and co-workers, *Sci. Am.*, 44 (Apr. 1983).

51. J. B. Price, in H. R. Huff and R. R. Burgess, eds., *Semiconductor Silicon*, The Electrochemical Society Proceedings Series, Electronics Electrothermics and Metallurgy Division, Electrochemical Society, New York, 1973.

52. D. L. Kendall, *Appl. Phys. Lett.* **26**, 195 (1975).

53. W. F. Clark, *IEEE Trans. Components, Hybrids, Manu.* **15**, 397 (1992).

54. S. Hanamura and co-workers, *IEEE Trans. Electron Dev.* **ED-34**, 94 (1987).

55. J. P. Colinge, *SOI Technology*, Kluwer Academic Publishers, Boston, Mass., 1991.

56. G. G. Shahidi and co-workers, *IEEE Trans. Electron Dev.* **41**, 2405 (1994).

57. G. G. Shahidi and co-workers, *IBM J. Res. Devel.* **39**, 229 (1995).

58. D. Vook and co-workers, *IEEE Trans Electron Dev.* **41**, 1013 (1994).

Kenneth Rose
Rensselaer Polytechnic Institute

AMORPHOUS SEMICONDUCTORS

Since 1970 the subject of amorphous semiconductors, in particular silicon, has progressed from obscurity to product commercialization such as flat-panel liquid crystal displays, linear sensor arrays for facsimile machines, inexpensive solar panels, electrophotography, etc. Many other applications are at the developmental stage such as nuclear particle detectors, medical imaging, spatial light modulators for optical computing, and switches in neural networks (1,2).

The success has been primarily due to the developments that occurred in the early 1970s (3) at the University of Dundee (United Kingdom) where it was demonstrated that a device-quality amorphous silicon semiconductor (*a*-Si) could be produced with the following features: low concentration of defects, high photosensitivity, ability to be doped, and no size limitation.

There is another class of amorphous semiconductors based on chalcogens which predate the developments that have occurred in *a*-Si. Because their use has been limited, eg, to switching types of devices and optical memories, this discussion is restricted to the optoelectronic properties of *a*-Si-based alloys and their role in some applications.

Classification

Amorphous semiconductors are solids in which the atoms retain well-defined local, or nearest neighbor, order but lack long-range periodic order of crystalline solids, and the local bonding between atoms is characterized predominantly by covalent forces (4,5). These types of semiconductors are normally synthesized as metastable thin solid films and are to be distinguished from glass which is formed by continuous hardening of a cooled liquid. Amorphous material is a solid in internal equilibrium in which there is, just as in crystalline solids, a definite set of equilibrium positions about which the atoms oscillate. However, in contrast to crystalline solids, the amorphous materials do not exhibit translational symmetry, ie, there is no long-range order.

The tetrahedrally bonded materials, such as Si and Ge, possess only positional disorder; however, materials of this type exhibit high density of defect states (DOS). It is only with the addition of elements such as hydrogen and or a halogen, typically fluorine, that the DOS is reduced to a point such that electronic device applications emerge. These materials contain up to ~10 atomic % hydrogen, commonly called hydrogenated amorphous silicon (*a*-Si:H).

Effect of Disorder

The short-range order in a material is important in determining optoelectronic properties. For instance, x-ray and electron diffraction experiments performed on amorphous silicon (*a*-Si) and germanium (*a*-Ge) have revealed that the nearest neighbor environments are approximately the same as those found in their crystalline counterparts (6); photoemission experiments performed on *a*-Si show that the DOS in valence and conduction bands are virtually identical to the corresponding crystal with the exception that the singularities (associated with periodicity) present in the latter are smeared out in the former.

Hence, the concept of DOS, $g(E)$, the number of one-electron states in the energy, E , range, $E + dE$, is valid along with the determination of average occupancy using Fermi-Dirac statistics. Another valid concept is mobility, μ (velocity per unit applied electric field), which depends on the entire state of the system. For a perfectly periodic crystal at absolute zero ($T = 0$ K), all the states are delocalized and the value of μ is infinite. In amorphous semiconductors, intense scattering, due to large fluctuations in the one-electron potential, can greatly reduce the value of μ . Once $g(E)$ and $\mu(E)$ are known, the optoelectronic properties of the material can be determined.

As is to be expected, inherent disorder has an effect on electronic and optical properties of amorphous semiconductors providing for distinct differences between them and the crystalline semiconductors. The inherent disorder provides for localized as well as nonlocalized states within the same band such that a critical energy, E_c , can be defined by distinguishing the two types of states (4). At $E = E_c$, the mean free path of the electron is on the order of the interatomic distance and the wave function fluctuates randomly such that the quantum number, k , is no longer valid. For $E < E_c$, the wave functions are localized and for $E > E_c$, they are nonlocalized. For $E > E_c$, the motion of the electron is diffusive and the extended state mobility is approximately $10 \text{ cm}^2 \text{ sV}$. For $E < E_c$, conduction takes place by hopping from one localized site to the next. Hence, at $E = E_c$, μ goes through a transition with a variation of about 10^3 . This leads to the concept of the mobility edge at $E = E_c$ and correspondingly at the valence band edge, $E = E_v$; $E_c - E_v$ is defined as the μ -gap which leads to the existence of semiconducting behavior in amorphous materials.

Growth of Amorphous Silicon

By far the most widely studied (7) and used deposition technique for preparing amorphous silicon is the plasma-enhanced chemical vapor deposition (PECVD) technique, which typically utilizes a capacitively coupled multichamber system, as shown in Figure 1. A gas, such as SiH_4 , is allowed to pass at a controlled rate between the two plates and the plasma is generated using d-c, audio, radio-frequency (r-f), or microwave frequencies. A variety of species are produced, such as atoms, free radicals, and stable and unstable ions. The mean energy of the electron is on the order of a few eV and electron temperature can be up to 100 times higher than that of the gas. Hence the electrons possess enough energy to break down the molecular bonds.

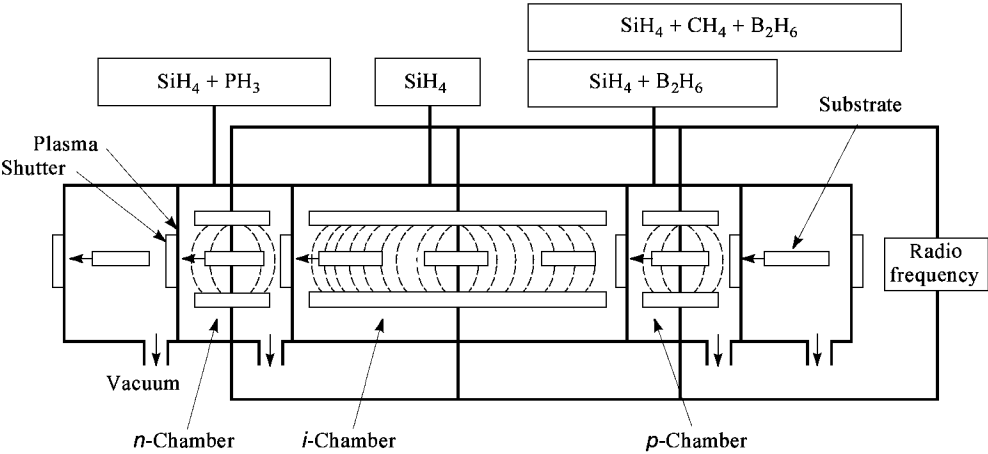


Fig. 1. Multichamber PECVD reaction chamber apparatus for fabrication of *a*-Si:H films, where the *i*-chamber represents the deposition of the intrinsic

layer; *n*-chamber, the deposition of *n*-type layer; and *p*-chamber, the deposition of *p*-type layer.

The defect state density within the μ-gap or the density of states (DOS) is critically controlled by the deposition temperature, *T*_p, of the substrates; the DOS decreases by several orders of magnitude when *T*_p is raised from 300 to 570 K (8). This change in DOS has been attributed to basic structural change from a polymeric-type (SiH_{*n*}) structure, for samples deposited at low temperature, to a monohydride (SiH) structure for higher *T*_p samples (9). The pressure of the gases during deposition also affects the properties of the film because gas-phase polymerization is encouraged at high pressures with the consequence that SiH₂ and SiH₃ grouping within the film becomes more pronounced and again DOS increases in the film.

The relative abundance of neutral SiH₄ and H₂ species have been measured as a function of power, pressure, flow rate, and dilution. For low power levels, eg, 5 W, up to 50° of the SiH₄ gas is dissociated and the percentage increases to 80° for a power of 50 W. The decomposition of SiH₄ gas proceeds more readily with lower flow rates. These observations, coupled with infrared (ir) measurements performed on the films, suggest that deposition under conditions in which the silane gas is not entirely decomposed leads to a majority of SiH units, whereas those deposited under conditions in which silane is strongly dissociated contain a majority of dihydride units leading to a deterioration of the semiconductor. Also, when the dwell time of SiH₄ in the plasma region increases, the resultant film exhibits a pronounced peak at 2090 cm⁻¹ from the ir spectra corresponding to SiH₂ inclusion.

Various plasma diagnostic techniques have been used to study the SiH₄ discharges and results have helped in the understanding of the growth kinetics. These processes can be categorized as r-f discharge electron kinetics, plasma chemistry including transport, and surface deposition kinetics.

The primary process of SiH₄ decomposition is electron impact which produces a large number of different neutral and ionic species as shown in Table 1. The density of SiH₂ and SiH₃ neutral species produced has been found to be much larger than the density of the ions. For example, mass spectrometric data for silane discharges indicate that the density of ionic species is lower by ~10⁴ compared with the density of neutral species. Further, mass spectrometer signals of ionic species, such as SiH⁺₃, SiH⁺₂, SiH⁺, SiH⁺, and Si₂H⁺₃, increase by more than two orders of magnitude as the r-f power is increased, eg, from 2 to 20 W. A rapid rise in the population of ions, with power, implicitly means an increase in electron density.

Table 1. Electron Impact Dissociative Processes Operative in Silane Plasma^a

Dissociation process	Enthalpy of reaction Δ <i>H</i> , eV
<i>e</i> + SiH ₄ → SiH ₂ + H ₂ + <i>e</i>	2.2
<i>e</i> + SiH ₄ → SiH ₃ + H + <i>e</i>	4.0
<i>e</i> + SiH ₄ → Si + 2 H ₂ + <i>e</i>	4.2
<i>e</i> + SiH ₄ → SiH + H ₂ + H + <i>e</i>	5.7
<i>e</i> + SiH ₄ → SiH ₂ ⁺ + H ₂ + 2 <i>e</i>	11.9
<i>e</i> + SiH ₄ → SiH ₃ ⁺ + H + 2 <i>e</i>	12.3
<i>e</i> + SiH ₄ → SiH ⁺ + H ₂ + H + <i>e</i>	8.7

^a Quantities Δ*H* are enthalpies of formation of the species; the densitiess of SiH₂ and SiH₃ species in the plasma have been shown to be larger than other species.

Table 2 shows the predominant reactions that take place during the transport of the species produced by electron impact dissociation to the substrate. The large rate of the insertion reaction of SiH₂ with SiH₄ results in the production of Si₂H₆ and is one of the mechanisms by which the powders observed using silane decomposition can be formed. The growth mechanism is complex because virtually all types of species reach the substrate surface. The exact mechanism as to whether the precursor for growth is primarily SiH₂ or SiH₃ is still controversial and further work needs to be performed to elucidate it.

Table 2. Reactions Within Plasma During Transport of Species Produced by Electron Impact Dissociation of Silane

Reaction	Enthalpy of reaction Δ <i>H</i> , eV
SiH ₄ + H → SiH ₃ + H ₂	0.5
SiH ₂ + SiH ₄ → Si ₂ H ₆	2.1
SiH ₂ + H ₂ → SiH ₄	2.2
SiH ₃ + SiH ₃ → Si ₂ H ₆	3.2
Si ₂ H ₆ → SiH ₂ + SiH ₄	2.1
SiH ₂ ⁺ + SiH ₄ → SiH ₃ + SiH ₃ ⁺	0.04
SiH ₃ ⁺ + SiH ₄ → Si ₂ H ₆ ⁺ + H ₂	0.12

The state-of-the-art *a*-Si:H films (Table 3) are deposited at the rate of 1–3 Å s with the gas utilization rate on the order of 15°. Larger gas utilization rates, hence larger deposition rates, usually result in inferior properties than those indicated in Table 3. Increasing the deposition rate by merely increasing the power leads to dust formation. The use of higher excitation frequency can lead to deposition rates in excess of 15 Å s and still give relatively good film properties (7).

Table 3. Typical Optoelectronic Parameters Obtained for *a*-Si:H(F) Alloys

Parameter	Value
<i>Undoped</i>	
hydrogen content, <i>C</i> _H , °	~10
dark conductivity at 300 K, σ _D , (Ωcm) ⁻¹	~10 ⁻¹⁰
activation energy, Δ <i>E</i> or <i>E</i> _a , eV	~0.8–0.9
pre-exponent conductivity factor, σ _o , (Ωcm) ⁻¹	>10 ³
optical band gap at 300 K, <i>E</i> _g , eV	1.7–1.8
<i>E</i> _g variation with temperature, <i>E</i> _g (<i>T</i>), eV K	2–4 × 10 ⁻⁴
density of states	
at minimum, <i>g</i> _{min} or <i>g</i> (<i>E</i> _D), cm ⁻³ eV ⁻¹	>10 ¹⁵ –10 ¹⁷
at conduction band edge, <i>g</i> (<i>E</i> _C), cm ⁻³ eV ⁻¹	~10 ²¹
electron spin resonance spin density, <i>N</i> _s , cm ⁻³	~10 ¹⁵

infrared spectra, cm ⁻¹	2000–640
photoluminescence peak at 77 K, eV	~1.25
extended state mobility	
electrons, μ _n or μ _p , cm ² (sV)	>10
holes, μ _p or μ _p , cm ² (sV)	~1
drift mobility	
electrons, μ _n or μ _p , cm ² (sV)	~1
holes, μ _p or μ _p , cm ² (sV)	~10 ⁻²
tail slope, meV	
conduction band	25
valence band	40
hole diffusion length, μm	~<0.5
<i>Doped amorphous</i>	
<i>n</i> -type ~1% addition PH ₃ to gas phase	
σ _Ⓐ (Ωcm) ⁻¹	~10 ⁻²
Δ <i>E</i> , eV	~0.2
<i>p</i> -type ~1% addition B ₂ H ₄ to gas phase	
σ _Ⓐ (Ωcm) ⁻¹	~10 ⁻³
Δ <i>E</i> , eV	~0.3
<i>Doped microcrystalline</i>	
<i>n</i> -type ~1% PH added to dilute SiH ₄ :H ₂ ^a gas mixture	
σ _Ⓐ (Ωcm) ⁻¹	<1
Δ <i>E</i> , eV	<0.05
<i>p</i> -type ~1% B ₂ H added to dilute SiH ₄ :H ₂ gas mixtures	
σ _Ⓐ (Ωcm) ⁻¹	<1
Δ <i>E</i> , eV	<0.05

^a Or 500 vppm PH₃ added to SiF₄:H₂ (8:1) gas mixture. Relatively high power is involved.

Properties

As discussed, an inclusion of an element such as H plays a crucial role in determining the properties of the *a*-Si:H material. This can be illustrated by considering the results of sputtered Si prepared in a reactive H environment. Figure 2 shows the effect of H on the electronic properties of *a*-Si with all the samples prepared at 200°C. In Figure 2 the dark conductivity, σ_D, is plotted as a function of 10³ / *T* with hydrogen pressure. The hydrogen content of the films increases from curves A to F. With the addition of 20 atomic % H to the films, curve F, an activated behavior over several orders of magnitude in the conductivity, results which corresponds to extended state conduction above *E*_g. This change in the conduction behavior is due to the reduced DOS within the μ-gap.

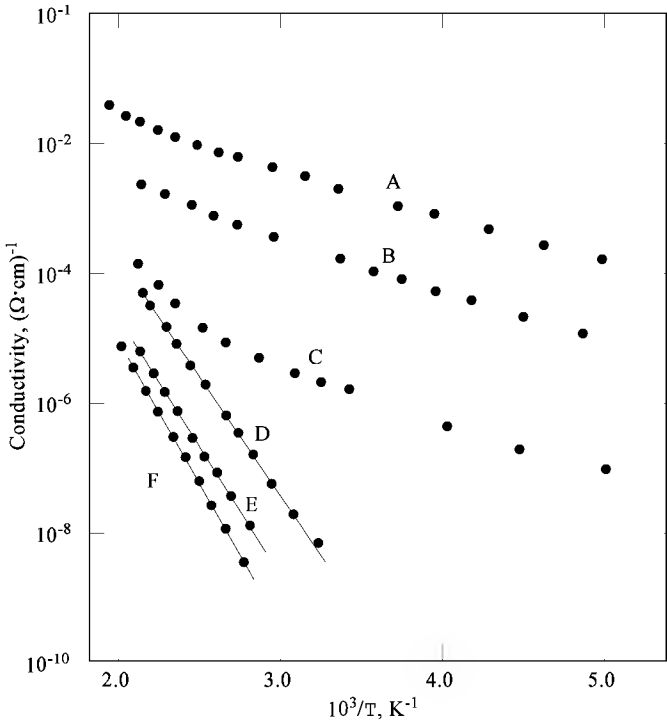


Fig. 2. Dark conductivity, σ_D, vs 10³ / *T* for sputtered *a*-Si:H samples prepared at *T*_r of 200°C using partial hydrogen pressure between 0–0.13 Pa (0–1 mtorr). Hydrogen content (atomic %) and hydrogen partial pressure are as follows for each curve: A, 0 atomic (at.) %; B, 1 at. %, 0.002 Pa; C, 2 at. %, 0.0067 Pa; D, 10 at. %, 0.02 Pa; E, 15 at. %, 0.05 Pa; and F, 20 at. %, 0.133 Pa. To convert Pa to mtorr, multiply by 7.5.

The optoelectronic properties of the *a*-Si:H films depend on many deposition parameters such as the pressure of the gas, flow rate, substrate temperature, power dissipation in the plasma, excitation frequency, anode–cathode distance, gas composition, and electrode configuration. Deposition conditions that are generally employed to produce device-quality hydrogenated amorphous Si (*a*-Si:H) are as follows: gas composition = 100% SiH₄; flow rate is high, ~40 cm³; pressure is low, 26–80 Pa (200–600 mtorr); deposition temperature = 250°C; radio-frequency power is low, <25 mW cm²; and the anode–cathode distance is ~1–4 cm.

Table 3 summarizes some of the present state-of-the-art parameters obtained for undoped and doped *a*-SiH(F) material thus produced. The device-quality material exhibits semiconductivity because $\ln \sigma$ vs $10^3/T$ exhibits a straight line with a conductivity activation energy of ~ 0.85 eV, which is approximately equal to half the band gap of ~ 1.75 eV. Typically the dark conductivity $< 10^{-10} (\Omega\text{cm})^{-1}$ changes to $> 10^{-5} (\Omega\text{cm})^{-1}$ under light illumination (Global AM1.5 sun illumination of intensity 100 mW cm^{-2}). The DOS (or defect) is found to be low with a dangling bond (DB) density, as measured by electron spin resonance (esr) of $\sim 10^{15} \text{ cm}^{-3}$. The inherent disorder possessed by these materials manifests itself as band tails which emanate from the conduction and valence bands and are characterized by exponential tails with an energy of 25 and 45 meV, respectively; the broader tail from the valence band provides for dispersive transport (shallow defect controlled) for holes with a low drift mobility of $10^{-2} \text{ cm}^2 (\text{sV})$, whereas electrons exhibit nondispersive transport behavior with a higher mobility of $\sim 1 \text{ cm}^2 (\text{sV})$. Hence the material exhibits poor minority (hole) carrier transport with a diffusion length $< 0.5 \mu\text{m}$, which puts a design limitation on electronic devices such as solar cells.

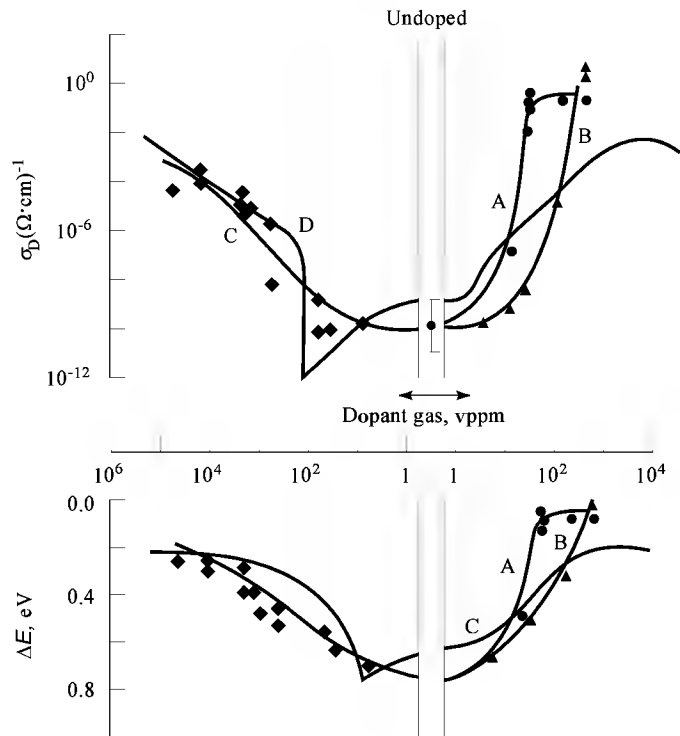


Fig. 3. The room temperature dark conductivity, $\sigma_D (\Omega\text{cm})^{-1}$, and conductivity activation energy, ΔE in eV, plotted as A, a function of vppm of AsH_3 (●); B, PH_3 (▲); and C, B_2H_6 (◆) into the premix gas ratio of $\text{SiF}_4\text{:H}_2 = 10 : 1$. The *p* to *n* transition (left to right) refers to *a*-Si:F:H alloy, and D refers to doping characteristics for this alloy.

The low DOS achieved in *a*-Si:H enables it to be readily doped, a prerequisite for any device application; *n*- and *p*-type doping is achieved by the addition of PH_3 and B_2H_6 to SiH_4 in the gas phase, respectively. Figure 3, a plot of σ_D and conductivity activation energy, ΔE , as a function of PH_3 and B_2H_6 content, shows that the most heavily *n*-type doping results in $\sigma_D \sim 10^{-5} (\Omega\text{cm})^{-1}$. By manipulating the plasma (using SiF_4 and H_2) or heavily diluting SiH_4 plasma with H_2 , microcrystalline *n*- and *p*-materials can be made with conductivities in excess of $1 (\Omega\text{cm})^{-1}$.

Another parameter of relevance to some device applications is the absorption characteristics of the films. Because the *k* quantum is no longer valid for amorphous semiconductors, *a*-Si:H exhibits a direct band gap (~ 1.70 eV) in contrast to the indirect band gap nature in crystalline Si. Therefore, *a*-Si:H possesses a high absorption coefficient such that to fully absorb the visible portion of the sun's spectrum only $1 \mu\text{m}$ is required in comparison with $> 100 \mu\text{m}$ for crystalline Si. Further improvements in the material are expected to result from a better understanding of the relationship between the processing conditions and the specific chemical reactions taking place in the plasma and at the surfaces which promote film growth.

Applications of *a*-Si-Based Semiconductors

As discussed previously, inherent disorder possessed by *a*-Si:H alloy limits the mobility of the free carriers (electrons and holes) to about $10 \text{ cm}^2 (\text{sV})$; this is compared with crystalline Si, in which the electron mobility is $1500 \text{ cm}^2 (\text{sV})$. However, crystalline Si is expensive to manufacture and its size is limited to about 20 cm in diameter. Many applications discussed have either emerged or been identified which preclude the use of crystalline Si because of cost, size, or both. The basic commonality in these applications is the ability to fabricate devices on areas much larger than can be addressed by crystalline Si. Furthermore, these applications are not demanding in terms of speed, which then provides *a*-Si:H alloy with a distinct competitive advantage.

Figure 4 shows the basic construction of the devices used in different applications, involving the deposition of multilayers of *a*-Si:H of intrinsic (*i*), doped (*n*⁺, *p*⁺), and closely allied films, such as amorphous silicon nitride, SiN, and transparent conducting oxide (TCO). As in crystalline semiconductors, the basic building block is a junction (*p-n*) or thin-film transistor. Attempts to fabricate *p-n* junctions (using *a*-Si:H) have thus far exhibited insignificant rectification, as shown in Figure 5. The reason is that during the process of doping with B or P, extra defect states are introduced within the gap, which leads to an increase in the generation–recombination current, thus precluding viable rectification. Instead of this type of junction, an *i*-layer is inserted between the *p*⁺ and *n*⁺ layers with the charge separation occurring within the intrinsic layer; a basic requirement being that the *p*⁺ and *n*⁺ layers should be thick enough to sustain a depletion width. Figure 5 also shows the forward and reverse bias characteristics of a *p-i-n* junction with a much improved rectification ratio greater than 10^5 at 0.8 V. The other type of structure is a field-effect transistor (FET) or thin-film transistor (TFT) and a typical construction is shown in Figure 4. In this case, the conductivity of the semiconducting layer is altered by the application of an external voltage applied to the gate metal electrode. For instance, an application of a positive gate voltage leads to a negative charge induced into the *a*-Si:H layer. The consequence is that the bands (valence and conduction) bend downward (accumulation layer) and conductivity (the current monitored between source and drain contacts) rises. An example of this is given in Figure 6, where it can be seen that with gate voltage of only a few volts, the source–drain current is altered by a factor of 10. These devices are used in some important areas, eg, solar cells, fax machines, and active matrix flat-panel displays.

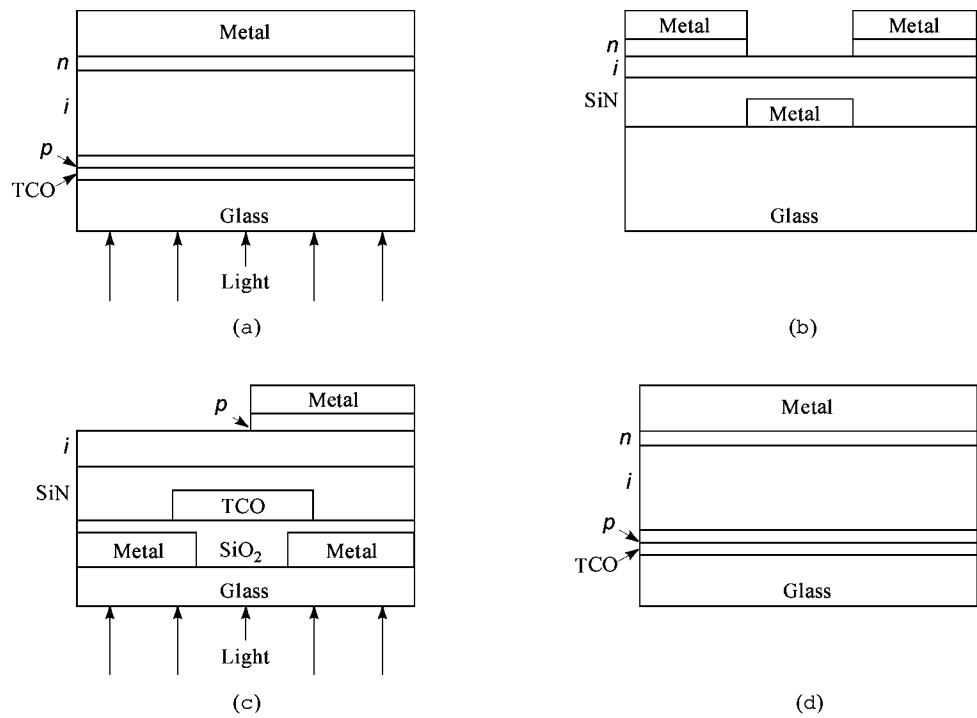


Fig. 4. Some electronic device applications using amorphous silicon: (a) solar cell, (b) thin-film transistor, (c) image sensor, and (d) nuclear particle detector. TCO = transparent conducting oxide.

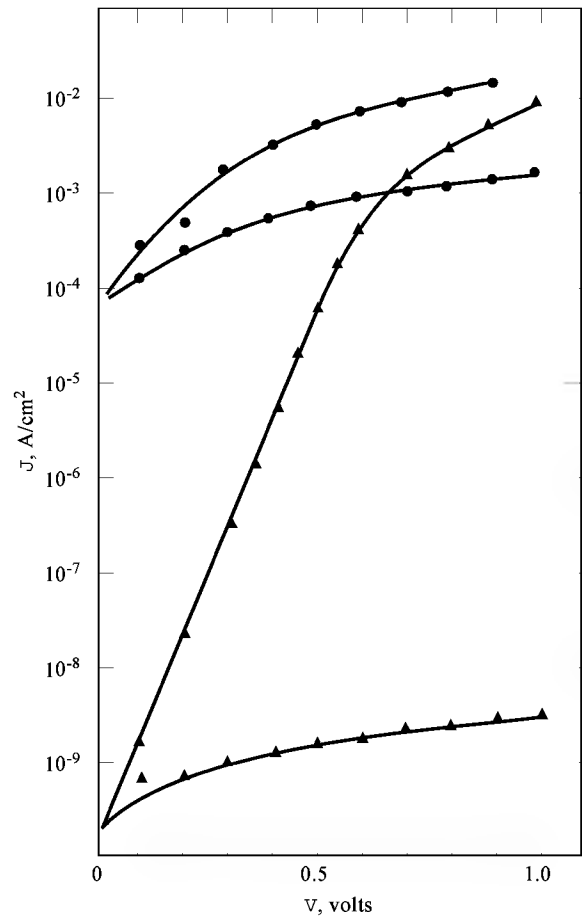


Fig. 5. Dark J - V (current-voltage) characteristics of p - n (●) and p - i - n (▲) junctions at 295 K.

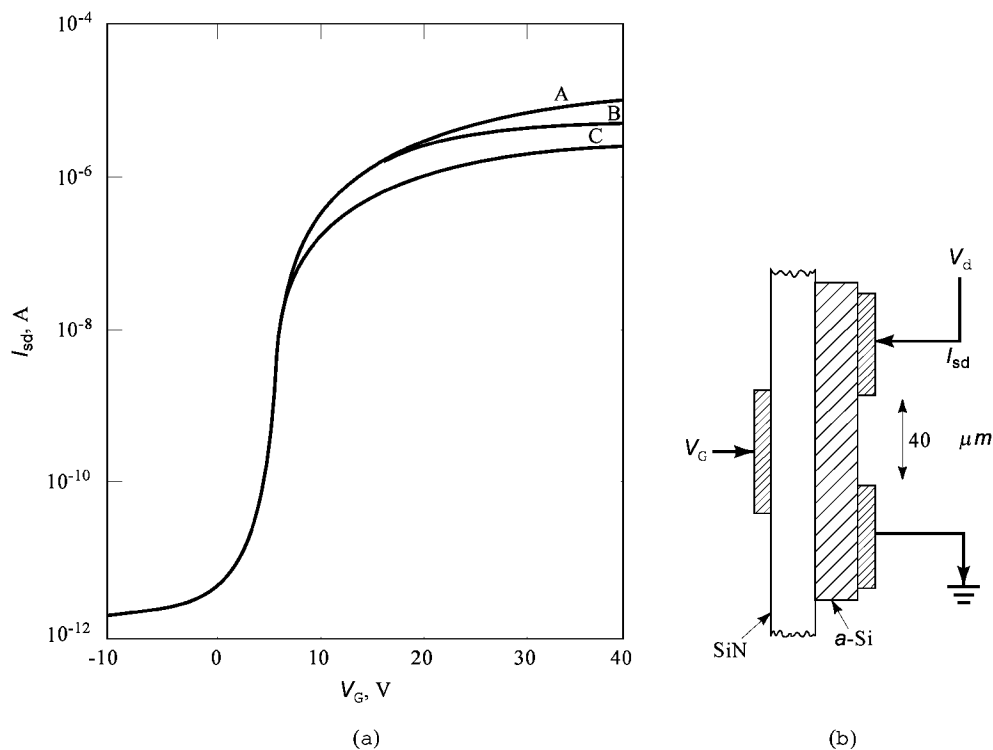


Fig. 6. (a) Transfer characteristics of (b) *a*-Si TFT element. The drain current, I_{sd} , is plotted against the gate voltage, V_G , for three drain potentials: A, $V_d = 20$ V; B, 10 V and C, 2 V.

The fact that excellent rectification can be demonstrated with these *p-i-n* junctions, together with high photosensitivity and the ability to deposit over large areas, leads to natural usage as a solar cell (see PHOTOVOLTAIC CELLS). The three primary parameters of a solar cell are the open-circuit voltage (V_{oc}), the short-circuit current (J_{sc}), and the fill factor (FF) with a theoretical maximum of 1.1 V, 22 mA cm², and 0.8, respectively, which leads to a conversion efficiency, η , of ~19%. The solar cell is generally constructed on a transparent conducting oxide, SnO₂ (see Fig. 4). A thin (10 nm) *a*-Si:C:B:H (*p*⁺ material) is deposited; in this, C is added to the *p*⁺ material during growth to widen the band gap and thus allow light to enter more effectively into the junction region primarily located at the *p-i* interface. After the *p*⁺ layer deposition, an *i*-layer of thickness between 300–700 nm is deposited, followed by an *n*⁺ layer (~20 nm) deposition, and the structure completed with a highly reflective contact, such as Ag. In order to minimize the losses, this structure is manipulated further with the insertion of extra layers, such as a graded layer between the *p*⁺ and *i* layer to minimize recombination losses that could exist at that discontinuity. Ag is often replaced with a two-layer reflective structure such as indium tin oxide / Ag. Generally, textured SnO₂ is used to promote multiple internal reflections to enhance the current produced by the device. Using some of these techniques, the efficiency of a single *p-i-n* junction in excess of 11% has been obtained. To further enhance the performance a concept of tandem junctions is employed. In this approach, the low energy photons are directed to solar cells made from smaller band gap material (eg, *a*-Si:Ge:H) and the high energy photons are directed to wide band gap cells where their energy is not dissipated by creating electron–hole pairs with energies much in excess of the band gap. The cells, whose active semiconductors possess different band gaps, are connected in series and stacked on top of each other with the widest band gap material facing the incident illumination. The low energy photons pass through the cells until they reach a band gap semiconductor that can utilize them. The thickness and values of the band gaps are arranged such that the current generated from the constituent cells are the same. Calculations leading to the design of multijunction stack arrays have been performed, indicating a maximum conversion efficiency of ~24% with stacks consisting of 1.45, 1.7, and 2.0 eV band gap materials. Experimentally, $\eta > 13\%$ has been achieved for a three-stacked cell on a stainless steel substrate, and $\eta > 10\%$ for large area modules on a glass substrate that exhibit <15% degradation over several hundred hours of constant testing (7).

The *p-i-n* junctions of virtually similar construction are also used as contact linear array image sensors in facsimile machines. An advantage is that they eliminate the need for magnification lenses which are used in conjunction with charge coupled device (CCD) sensors, constructed from crystalline Si devices. In a facsimile machine, the sensor unit consists of an illuminator, a compact optical guide, and a long photosensor array connected to the scanner. Light from the light-emitting diode (LED) arrays illuminates the document and the reflected signal is guided by the rod lens array onto the *a*-Si:H *p-i-n* junction photosensor array (see Fig. 4). The *p-i-n* junctions are constructed in a pixelized fashion with the number dictated by the resolution requirements, eg, 200 dots per inch (80 dots / cm) necessitates an excess of 1800 pixels for a page width array. The junctions are operated in reverse bias, with typical off-current of ~10⁻¹³ amps, and under illumination the current rises by several orders of magnitude to excess of 10⁻⁹ amps. The photocurrent generated is proportional to the light intensity reaching the pixels. By operating the devices in the charge collection mode and with use of appropriate circuitry, the output of the device is proportional to light intensity. The elements are attached to read-out integrated circuits which then transmit the data to the receiver.

In an active matrix liquid crystal (LC) display application, a single TFT for each picture element (pixel) is incorporated, as shown in Figure 7. In a display of 12 in. (~30 cm) diagonal size, pixels in excess of 10⁶ have been incorporated. The display consists of two plates between which the LC is sandwiched and backlit. The LC possesses the property that when its twist is altered, the transmission characteristics are changed. $S_1 \dots S_n$ and $G_1 \dots G_n$ are bus lines for the source and gate electrodes, and the drain, *D*, contact is connected to each TCO square (see Fig. 7). The LC is in series with the drain circuit and behaves electrically as a capacitor, C_{LC} . The addressing scheme involves writing one line at a time, corresponding to a television line, in 63.5 μs. In this time scale, gate G_1 is simultaneously applied to the source lines, ie, zero potential for pixels that are to be turned on. After 63.5 μs, the TFTs that have been addressed (G_1) are turned off and the charge is retained in C_{LC} . In the next 63.5 μs, video data are entered into G_2 , and the information for that is simultaneously applied to the source lines. When all gate buses have been addressed, the scan returns to the first bus line and the process is repeated.

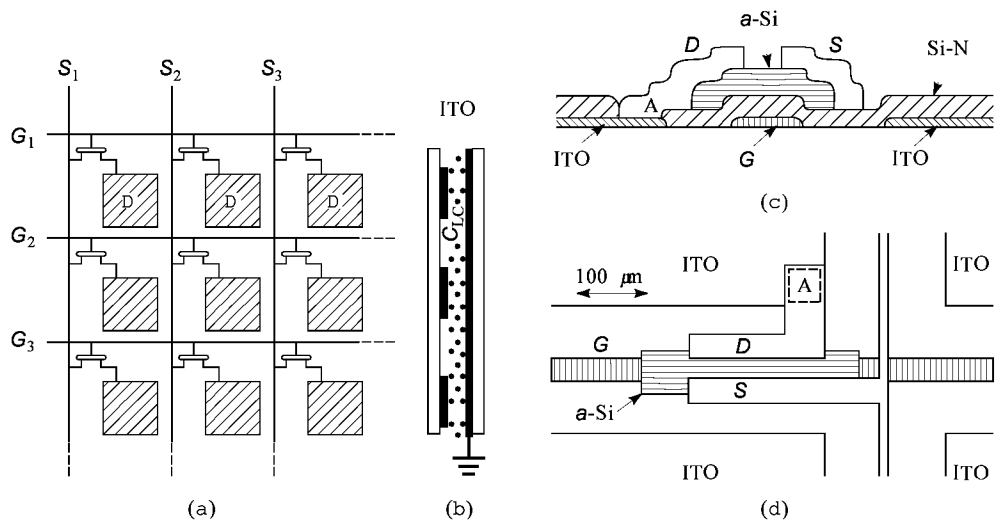


Fig. 7. Schematic layout of addressable liquid crystal (LC) panel: (a) plane view, (b) side view, (c) section through transistor element, and (d) TFT in part of matrix array. ITO – indium tin oxide.

Stability of Amorphous Silicon

Stability is one of the principal issues confronting the eventual commercialization of solar cells for large-scale power applications. There has been an increasing amount of data showing that good stability can be attained in *a*-Si alloy devices. In a solar cell device configuration, the degradation is seen primarily as a decrease in the fill factor and J_{sc} , but the device can be restored to its original performance level upon annealing at 175°C for 30 minutes. The amount of degradation is a function of the operating conditions: it is greatest at V_{oc} , reduces at J_{sc} , and is virtually eliminated under far reverse bias conditions. These observations have been associated with the so-called Staebler-Wronski effect (SW) where the application of light can significantly degrade the electronic properties of the material. Two types of states can be distinguished: a fully annealed state attained by slowly cooling the material from approximately 500 K to room temperature in the dark (state A); and a light-soaked state attained after application of $\sim 10^{21}$ photons cm^{-2} to the material (state B).

There have been many investigations of photoinduced effects in *a*-Si:H films linked to material parameters. Changes have been observed in the carrier diffusion length, unpaired spin density, density of states in the gap, and infrared transmission. The transition from state A to B seems to be induced by any process that creates free carriers, including x-ray radiation and injection (double) from the electrodes. Because degradation in a solar cell is accentuated at the open-circuit voltage conditions, the A to B transition occurs upon recombination of excess free carriers in which the energy involved is less than the band gap. It has been pointed out that this transition is a relatively inefficient one and the increase in spin density takes place at a rate of 10^{-8} spins per absorbed photon.

The recovery from state B to A seems to be complete upon annealing, suggesting that state B is metastable and that the photoinduced defects are distinct from the defects present in state A. Various explanations for the SW effect have been put forth but no clear consensus has yet emerged. Despite these problems the degradation phenomena can perhaps be engineered out for a solar cell. The degradation may be self-limiting because the number of created dangling bonds saturate, which in turn means that the collection width decreases and eventually saturates to a smaller value. Therefore, if the device thickness is approximately equal to the smallest collection width attained, the device should exhibit relative stability and the problem of degradation could be largely circumvented. Attempts at this involve use of tandem junctions in which very thin films (<250 nm) are used and the results indicate an improved stability.

Economic Aspects

Industry based on the use of amorphous silicon represents a multibillion dollar market and is growing at a rapid rate. Major sales are in the display area with numerous companies, eg, Toshiba, Sharp, and Epson involved; in the electrophotography arena, the companies which dominate are Canon, Minolta, and Sharp. Although solar energy attracts the most attention, it is in fact quite a small revenue earner for the *a*-Si-based industry, as it commands approximately 30% of the annual market of 70 MW yr with an average sale price of about \$4 W; the primary companies producing these products are Sanyo (Japan), Neste (France), Sinonar Corporation (Taiwan), and Solarex (U.S.). Nevertheless, the use of solar panels is expanding at rate of 20–30% yr and the consensus is that thin-film silicon is likely to dominate energy production, using photovoltaics, as this is likely to be the cost-effective solution to energy needs.

BIBLIOGRAPHY

"Semiconductors (Amorphous)" in *ECT* 3rd ed., Vol. 20, pp. 654–673, by D. Adler, Massachusetts Institute of Technology.

1. M. Hack, E. A. Schiff, A. Madan, M. Powell, and A. Matsuda, eds., "Amorphous Silicon Technology", *MRS Proc.* **377** (1995).
2. A. Madan and M. Shaw, *The Physics and Applications of Amorphous Semiconductors*, Academic Press, Inc., San Diego, Calif., 1988.
3. W. E. Spear and P. G. LeComber, *J. Noncryst Solids* **727**, 8–10 (1972); A. Madan, Ph.D dissertation, University of Dundee, U.K., 1973; W. E. Spear and P. G. LeComber, *Phil. Mag.* **33**, 935 (1976).
4. N. F. Mott and E. A. Davis, *Electronic Transport in Non-Crystalline Materials*, Clarendon Press, Oxford, U.K., 1979.
5. R. Zallen, *Physics of Amorphous Solids*, John Wiley & Sons, Inc., New York, 1983.
6. S. C. Moss and J. F. Graczyk, *Phys. Rev. Lett.* **23**, 1167 (1969).
7. G. Bruno, P. Capezzuto, and A. Madan, eds., *Plasma Deposition of Amorphous Silicon Based Materials*, Academic Press, Inc., New York, 1995.
8. A. Madan, P. G. LeComber, and W. E. Spear, *J. Noncryst. Solids* **20**, 239 (1976).
9. G. Lucovsky, R. Nemanich, and J. Knights, *Phys. Rev.* **B19**, 2064 (1979).

Arun Madan
MVSystems, Inc.



COMPOUND SEMICONDUCTORS

In the most general sense, the term compound semiconductor encompasses a large number of materials, most of which crystallize in either the zincblende (cubic), wurtzite (hexagonal), or rock salt (cubic) crystal structures. Prototypes of these structures are GaAs, CdS, and PbS, respectively. More exotic examples of zincblende semiconductors are ZnSiP₂ and Ga₂Se₃. The latter crystal is a defect zincblende structure in which one-third of the Ga sites are structural vacancies. Groups 13 and 15 (III–V) and (2, 12) and 16 (II–VI) compounds which crystallize in the zincblende structure are of significant value for electronic and photonic device applications. The III–V nitrides, AlN, GaN, and InN, crystallize in the wurtzite structure and are being developed for blue light-emitting devices and high temperature and high power electronics (1,2). The Group (2, 12) and 14 (II–VI) compounds are of value primarily for photonic applications. Blue lasers have been created from ZnMgSSe alloy heterostructures (2,3). The HgCdTe alloy system has been extensively studied for infrared photovoltaic detectors (4,5). CdTe has a band gap of 1.49 eV, whereas HgTe is a semimetal. Consequently, the band gap of the alloy may be varied smoothly from over 1 eV to zero. Another interesting class of materials are the dilute magnetic semiconductors (6). The transition-metal Mn, with a half-filled *d* shell (3*d*⁵), may be readily alloyed with all of the simple II–VI semiconductors. These materials have applications in flat-panel displays and are projected to apply to mid-infrared detectors and nonreciprocal optical devices (isolators and circulators). The Group 14–16 (IV–VI) lead salt crystals have a unique niche in mid-infrared detectors and lasers.

This article focuses primarily on the properties of the most extensively studied III–V and II–VI compound semiconductors and is presented in five sections: (1) a brief summary of the physical (mechanical and electrical) properties of the zincblende cubic semiconductors; (2) a description of the metal organic chemical vapor deposition (MOCVD) process. MOCVD is the preferred technology for the commercial growth of most heteroepitaxial semiconductor material; (3) the physics and (4) applications of electronic and photonic devices; and (5) the fabrication process technology in use to create both electronic and photonic devices and circuits.

Physical Properties

The defining characteristic of a semiconductor is that there is a gap between the completely filled valence states and the lowest excited or conduction states of the crystal which is typically <2.5 eV. A schematic band structure for zincblende cubic direct gap semiconductors is shown in Figure 1. The three lowest bands are valence, *v*, bands and the upper three minima derive from the lowest conduction, *c*, band. The features of relevance to the design and performance of electronic and photonic devices are the valence band maximum, Γ_v , and the conduction band minima at the symmetry points Γ_c , L, and X. GaP has its lowest conduction band minimum at X; in GaAs it is at Γ_c . The ordering of the conduction bands in GaP is similar to that of Si. The wide application of compound semiconductors is, however, due to the type of conduction band structure associated with GaAs. Because the minimum band gap is $\Gamma_c-\Gamma_v$, and noting that the momentum of a photon is negligible, light emission via the direct, momentum conserving recombination of electrons (in the conduction band) with holes (in the valence band) is possible. In indirect semiconductors such as Si and GaP light emission is not readily obtained from recombination across the minimum band gap because momentum is not conserved in a simple photon emission process. For this reason virtually all light-emitting diodes (LEDs) and diode lasers are fabricated from compound semiconductors.

The primary advantage of a direct gap semiconductor is its ability to strongly absorb and emit light at the fundamental band gap. The secondary advantage is that a small isotropic electron-effective mass is typically associated with the Γ_c minimum. The effective mass of an electron or hole in a crystal is inversely proportional to the curvature of the band in which it resides. In all zincblende semiconductors the effective mass at Γ_c is nearly isotropic and universally less than the average of the anisotropic mass associated with either the L or X minima. It is also approximately proportional to the magnitude of the band gap. In GaAs the electron effective mass is 0.067 *m*₀ where *m*₀ is the free-electron mass. This may be compared to the electron conductivity effective mass of Si which is 0.260 *m*₀. The effect of band structure on transport is illustrated by the electron velocity vs electric field curves for GaAs and Si in Figure 2. For low electric fields the steeper slope for GaAs implies an electron mobility roughly six times greater than in Si. This implies shorter switching transients at lower voltages in field-effect transistors (FETs). At high electric fields, electron transport in GaAs is limited by carrier scattering into the low mobility L minima and much of the speed advantage is lost. However, because the saturated velocity is obtained at much lower fields there is an advantage that GaAs FETs can operate at the same speed as a Si FET while consuming less power. This is significant for high density integrated circuits.

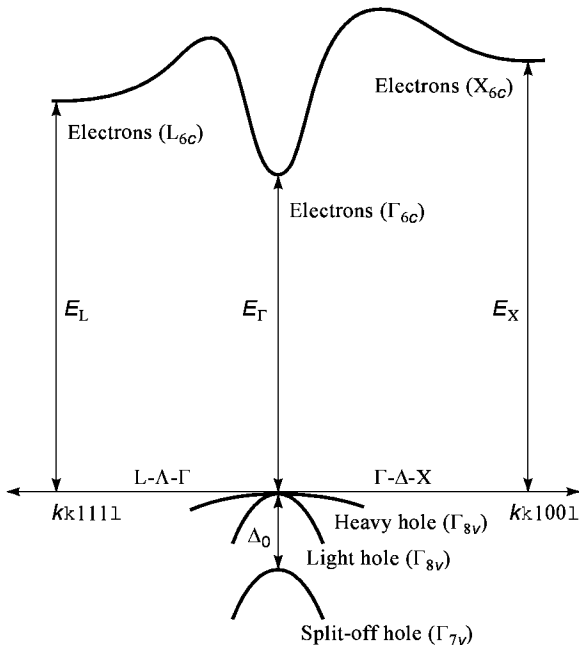


Fig. 1. Representation of the band structure of GaAs, a prototypical direct band gap semiconductor. Electron energy, *E*, is usually measured in electron volts relative to the valence, *v*, band maximum which is used as the zero reference. Crystal momentum, *k*, is in the first Brillouin zone in units of 2π a

where a is the lattice constant. In semiconductors with an indirect band gap the ordering of the valence (hole) states is unchanged, whereas the ordering of the conduction, c , states is inverted to $X < L < \Gamma$. In the semimetal HgTe the conduction band lies between the Γ_{8v} and Γ_{7v} valence bands and the curvatures of the Γ_{8v} light hole band and the Γ_c electron band are inverted with respect to those shown for GaAs. Δ represents the line of symmetry.

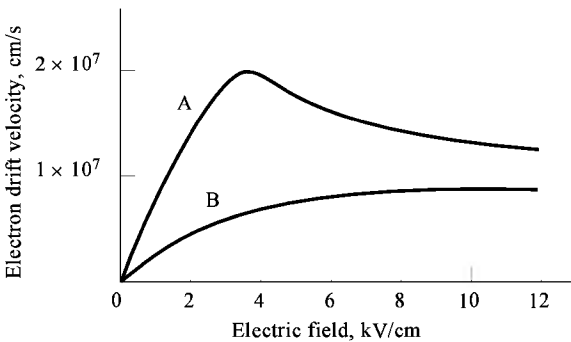


Fig. 2. Electron drift velocities as a function of electric field for A, GaAs and B, Si. The gradual saturation of curve B is characteristic of all indirect semiconductors. Curve A is characteristic of direct gap semiconductors and at low electric fields this curve has a steeper slope which reflects the larger electron mobility. The peak in curve A is the point at which a substantial fraction of the electrons have gained sufficient energy to populate the indirect L minimum which has a much larger electron-effective mass than the Γ minimum. Above 30 kV cm (not shown) the drift velocity in Si exceeds that in GaAs.

The tertiary advantage of compound semiconductors is the ability to fabricate multilayer structures epitaxially, widely varying the material properties such as band gap and refractive index from layer to layer. Within the constraint that the lattice constants are well matched it is usually possible to grow heterostructures of different III–V and II–VI semiconductors with atomically abrupt interfaces and monolayer thickness control. With some intermixing at the interfaces it is also possible to grow lattice matched heterostructures between elemental, III–V, and II–VI semiconductors. The possible combinations are essentially infinite and combining different alloy layers of variable electronic and optical properties allows for the design of a wide variety of complex devices, many of which are explicitly designed to exploit quantum size effects. The prevalent growth technologies are metal organic chemical vapor deposition (MOCVD) and molecular beam epitaxy (MBE) (7,8). In MOCVD, deposition takes place on a heated substrate with the constituent elements transported as alkyls, in a hydrogen carrier gas, and hydrides. In MBE, deposition occurs in an ultrahigh vacuum by evaporating elemental beams from pyrolytic boron nitride crucibles that are resistively heated. Other commercial crystal growth technologies which are less flexible than MOCVD and MBE are hydride vapor-phase epitaxy (VPE) and liquid-phase epitaxy (LPE). These techniques are limited either in the specific materials which can be grown, in their ability to grow thin layers with abrupt junctions, or both.

Table 1 lists the properties of several semiconductors relevant to device design and epitaxy. The properties are appropriate to the zincblende crystal structure in those cases where hexagonal polytypes exist, ie, ZnS and ZnSe. This first group of crystal parameters applies to the growth of epitaxial heterostructures; the cubic lattice constant, a , the elastic constants, c_{11} , c_{12} , and c_{44} ; and the congruent sublimation temperature, T_s . For growth of defect-free heterostructures either a uniform lattice constant must be enforced or lattice mismatched layers must be constrained to be thin enough to allow purely elastic deformation which is thermodynamically stable against plastic deformation.

In the design of heteroepitaxial layers, which are not lattice matched, the elastic constants of the layers are essential to the calculation of the thermodynamically stable maximum layer thickness for elastic deformation as well as the shape of the deformed unit cell. In most cases the $\langle 001 \rangle$ surface is the preferred surface for crystal growth. In this case the deformation is purely tetragonal and expressions for the critical layer thickness and the dimensions of the unit cell are readily obtained (9). For growth on substrates with an arbitrary surface orientation the results are more difficult to obtain. An interesting case of lattice mismatched epitaxy is the case of a strained layer superlattice in which elastically deformed layers under compressive stress alternate with layers under tensile stress. By designing the magnitude of the tensile and compressive stresses to be equal, strained layer superlattices of arbitrary thickness may be grown with zero net stress which are stable against plastic relaxation (10).

Table 1. Physical Properties of the Cubic Zincblende Structure III–V and II–VI Semiconductors

Compound	a , ^a nm	c_{11} , ^b GPa ^c	c_{12} , ^b GPa ^c	c_{44} , ^b GPa ^c	T_s , ^d °C
III–V					
AlP	0.56435	(136)	(67)	(63)	
AlAs	0.56612	116.3	57.6	54.1	850
AlSb	0.61356	83.1	41.1	38.6	525
GaP	0.54506	141.4	64.0	70.3	700
GaAs	0.56535	119.0	53.8	59.4	640
GaSb	0.60955	88.3	40.2	43.2	475
InP	0.58688	101.1	56.1	46.5	540
InAs	0.60585	86.6	48.5	39.5	500
InSb	0.64789	66.6	37.6	31.1	400
II–VI					
ZnS	0.54102	98.1	62.7	44.83	
ZnSe	0.56676	81.0	48.8	44.1	
ZnTe	0.61037	71.3	40.7	31.2	
CdTe	0.6481	53.3	36.5	20.44	
HgTe	0.6401	53.61	36.60	21.23	

^a In the zincblende structure the bond length is related to the cubic lattice parameter as $(3a/4)^{-1}$.
^b The C_{ij} are the elastic constants; the bulk modulus of the material is computed as $B = (c_{11} + 2c_{12})/3$. Values in parentheses are estimates.
^c To convert GPa to dynes cm², multiply by 10^{10} .
^d The maximum congruent sublimation temperature, T_s , corresponds to the temperature above which the anion evaporation rate exceeds the cation evaporation rate from a $\langle 001 \rangle$ crystal surface in a vacuum.

Along with the lattice constant the congruent sublimation temperature is a useful guide as to which pseudobinary alloys, $(AC)_x(BC)_{1-x}$ or $(AC)_x(AD)_{1-x}$, may be easily grown. Similar arguments apply to pseudoternary and pseudoquaternary alloys. When the lattice constants of the end-point binaries are closely matched the alloy is easily obtained for all mole fractions. Large lattice mismatches typically imply a miscibility gap in the alloy phase diagram. A large difference in the congruent evaporation temperatures also implies difficulty in growing high quality material. In the MBE process the highest quality films are typically obtained by growing at temperatures just below T_c with a nearly stoichiometric arrival rate of cation and anion fluxes. Growth of alloy films is controlled to a large extent by the binary constituent with the lower T_c . A growth temperature near the lower binary T_c is required to ensure that all of the incident cation atoms are incorporated. On the other hand a low growth temperature relative to the largest binary T_c enhances the formation of intrinsic point defects associated with the high T_c component of the pseudobinary alloy. Native defects usually have a deleterious effect on both the electrical and optical properties of the semiconductor.

The remaining parameters are useful in the design of electronic and photonic devices. The positions of the three lowest conduction band minima, E_{Γ} , E_L , and E_X , are given relative to the valence band maximum, which is universally at the center of the Brillouin zone, Γ . The valence band maximum is composed primarily of p symmetry states derived from the anion. Spin-orbit interactions split what would be a threefold degenerate critical point into a twofold degenerate upper band and a nondegenerate lower band. Away from Γ the upper band splits into light and heavy hole bands. Neither band has spherical constant energy surfaces and three Luttinger parameters, γ_1 , γ_2 , and γ_3 , are required to describe the hole effective masses, m^* , in an arbitrary direction away from Γ . The relation is as follows where m_0 is the free-electron mass, k is momentum,

$$m_0 \ m_{ll(hh)}^*(k) = \gamma_1 \pm \left[4\gamma_2^2 + 12(\gamma_3^2 - \gamma_2^2) (k_x^2 k_y^2 + k_y^2 k_z^2 + k_x^2 k_z^2) \right]^{1/2}$$

and k is a unit vector in momentum space. The plus sign refers to the light holes, lh , and the minus sign to the heavy holes (hh) (Table 2). For most calculations the mass of the spin-orbit split-off band is unimportant and the light and heavy hole bands are approximated by directionally averaged parabolic bands. In many practical applications (11,12) this is accomplished by defining the parameter $\mu = (6\gamma_3 + 4\gamma_2) / 5\gamma_1$. In this approximation,

$$m_0 \ m_{ll(hh)}^* = \gamma_1(1 \pm \mu)$$

Table 2. Transport Properties^a of the Cubic Binary Compound Semiconductors

Compound	m_{Γ}	γ_1^b	γ_2^b	γ_3^b	m_{ρ}^c	$m_{\theta\theta}^c$
III–V						
AlP	(0.169)	(3.00)	(0.63)	(1.15)	(0.205)	(0.896)
AlAs	0.146	3.46	0.88	1.26	0.176	0.804
AlSb	0.12	4.69	1.24	1.62	0.131	0.570
GaP	0.172	4.04	0.53	1.26	0.167	0.475
GaAs	0.067	6.85	2.10	2.90	0.083	0.592
GaSb	0.041	13.30	4.20	5.50	0.043	0.299
InP	0.079	4.59	1.65	2.35	0.110	1.238
InAs	0.023	20.4	8.30	9.10	0.026	0.352
InSb	0.014	33.20	14.80	15.50	0.016	0.362
II–VI						
ZnS	(0.295)	(2.31)	(0.34)	(0.65)	(0.309)	(0.724)
ZnSe	0.165	3.0	0.53	0.92	0.229	0.614
ZnTe	0.130	3.9	0.83	1.30	0.168	0.536
CdTe	0.096	4.11	1.08	1.95	0.144	0.797
HgTe ^d	0.033	15.6	9.6	11.2	0.028	0.205

^a Values in parentheses are estimates.
^b Luttinger parameters.
^c The light (l) and heavy (h) hole masses are the spherical average values defined in text.
^d The negative electron and light hole masses listed for HgTe are a consequence of its being a semimetal rather than semiconductor. The curvatures of these two bands are inverted with respect to the convention defined for semiconductors.

The relevance of the indirect minima E_L and E_X are in determining the transport properties of the material. For all of the Al compounds and GaP, the lowest conduction band is at X and the valence band maximum is at Γ . The other bands follow in the order $E_L > E_{\Gamma}$. This is the same relative ordering found in Si. The X minima may be approximately described by constant energy surfaces which are ellipsoids of revolution about the line of symmetry (denoted Δ) connecting Γ to X. Along Δ the longitudinal effective mass, m_p is fairly large, on the order of 0.9–1.2 m_0 . The transverse effective mass, m_{ρ} in the plane perpendicular to Δ is usually $\sim 0.25 \ m_0$. The conductivity effective mass is obtained as $m_c^* = 3 \ (m_l + 2 \ m_t) \approx 0.33 \ m_0$. This is substantially larger than the isotropic Γ point effective masses obtained in the direct gap III–V semiconductors, where $m_{\Gamma} = m_c^*$, which range from 0.014 m_0 for InSb to 0.079 m_0 for InP. It is because of the smaller effective mass in direct gap semiconductors that they are superior for the fabrication of high speed low power transistors.

This advantage over indirect semiconductors, including Si, diminishes as improvements in processing technology continue to shrink device dimensions. In long devices transport is dominated by the carrier mobility which is inversely proportional to m^* . In short devices performance is limited by the saturated drift velocity. Short devices imply large electric fields and electrons accelerated by the field are scattered into higher minima, primarily E_L , where they have a larger m^* . In this limit the transport of electrons in direct and indirect semiconductors is very similar. In general, the performance advantage of a direct gap semiconductor in small device applications improves as the separations, $E_L - E_{\Gamma}$ and $E_X - E_{\Gamma}$, increase. This criterion would seem to imply that semiconductors such as InAs and InSb should make good transistors. There is a trade-off in that with decreasing E_{Γ} the electrons accelerated by the electric field quickly gain enough energy to interact with the lattice and break a bond releasing an electron–hole pair into the conduction channel of the device. This can lead to a runaway process known as avalanche multiplication which kills transistor performance. Most commercial transistors are made from GaAs, InP, and alloys with similarly large direct band gaps (Table 3).

An extensive compilation of the properties of compound semiconductors may be found in the Landolt-Bornstein reference books (13,14). Various subvolumes in the series cover the properties of elemental, III–V, II–V, and other less common semiconductors. Information may also be found concerning semiconductor technology. Another useful source of information is the EMIS data review series (15). These books describe the properties and technology of GaAs, HgCdTe, InP, AlGaAs, InGaAs, and the III–V nitride compounds.

Metal Organic Chemical Vapor Deposition

Metal organic chemical vapor deposition (MOCVD) is a technique that has evolved since 1975 into one of the two prime growth methods for compound semiconductor devices. This is because of the capability to grow high quality, very thin layers of most III–V and II–VI semiconductors making it the most versatile of the growth methods in use. Many comprehensive reviews have been written on this technique (7,16,17). MOCVD is also referred to as organometallic vapor-phase epitaxy (OMVPE) as well as MOVPE and OMCVD. Epitaxy is the term used to describe the growth by chemical reaction of a thin layer of semiconductor on the surface of a bulk crystal (wafer) of similar or different material, where the layer has a well-defined crystalline orientation with respect to the bulk crystal. MOCVD growth typically involves decomposition in a cold wall reactor of chemical precursors, usually an organometallic Group III compound and a Group V hydride. For the growth of gallium arsenide [1303-00-0] from trimethylgallium and arsine a typical reaction is as follows:

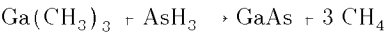


Table 3. Band Gaps and Dielectric Properties^a of Cubic Binary Compound Semiconductors at RT

Compound	E_F , eV	E_L , eV	E_X , eV	Δ_0 , eV	ϵ_0	n_∞
III–V						
AlP	3.60		2.45	(0.053)	9.40	2.689
AlAs	3.03	2.36	2.13	0.302	10.18	2.857
AlSb	2.30	2.21	1.62	0.673	11.86	3.228
GaP	2.78	2.64	2.39	0.085	11.00	3.011
GaAs	1.42	1.74	1.19	0.341	12.84	3.298
GaSb	0.72	0.80	1.03	0.764	14.73	3.689
InP	1.34	2.03	2.38	0.108	12.35	3.094
InAs	0.35	1.55	1.90	0.371	14.28	3.451
InSb	0.18	1.35	1.79	0.803	17.70	3.965
II–VI						
ZnS	3.74			0.067	8.9	2.39
ZnSe	2.67			0.42	7.6	2.32
ZnTe	2.28			0.97	7.28	3.18
CdTe	1.490			0.81	7.4	3.19
HgTe	0.303 ^b			1.08	21.0	3.90

^a Values in parentheses are estimates. Values for band gaps decrease with increasing temperature, whereas values for the static dielectric constant and long wavelength refractive index increase with increasing temperature.
^b Measured at 4 K.

This technique has been extended to all of the important III–V semiconductors such as AlGaAs, InP, InGaAs, and InGaAsP. Semiconductors grown by MOCVD for use as lasers, solar cells, detectors, and high speed transistors are in commercial production. MOCVD is a state-of-the-art technique for preparing epitaxial layers of compound semiconductors. It is a rapidly growing area as reflected by the number of papers and attendees at the biannual international conference. In 1981, 40 papers were presented at the international conference which had 112 attendees. In 1990, 199 papers were presented at the conference which was attended by 472 people. The conference has since had to limit the number of papers and attendees. Topics of interest to researchers in MOCVD include the synthesis and use of novel sources, the continued improvement of the materials' purity, understanding of the parameters that limit the sharpness of interfaces in heterostructures, the growth of quantum wires and dots, the reaction kinetics and growth mechanisms of MOCVD, ordered structures and how to control the ordering, the use of *in situ* diagnostics such as x-ray diffraction and optical reflection to control and better understand the MOCVD process, new reactor designs for multiwafer commercial growth, and extension of the technique to the growth of nonsemiconductors such as superconducting and ferroelectric oxide films (18).

CVD. Chemical vapor deposition is usually defined as a technique that involves a chemical reaction which results in the deposition of a solid on a specific surface (19). Growth techniques such as MOCVD, chloride or hydride vapor-phase epitaxy, and gas source molecular beam epitaxy are considered to be CVD techniques. Growth methods such as liquid-phase epitaxy, molecular beam epitaxy, and sputtering are not considered CVD techniques because they lack one or more of the criteria in the definition. CVD offers the advantages of near equilibrium growth which allows for the precise control of growth rates and improved purity. The use of CVD allows one to utilize a growth pressure from atmospheric down to very low pressures ($<1.3 \times 10^{-3}$ Pa) as well as a large variety of reactants with widely varying chemical reactivities. The capability for using materials that would decompose in the liquid or high temperature gaseous state is also an advantage of CVD. MOCVD has demonstrated very precise and reproducible control of alloy composition, layer thickness, and dopant levels. CVD techniques are generally readily scaled-up for mass production. Some disadvantages of the MOCVD technique include a large number of control variables, reactive atmospheres which can attack substrates and the growth apparatus, and the complexity of the apparatus, the mass transport, and the chemical reactions.

A complete understanding of the MOCVD process at a fundamental level would necessitate knowing the thermodynamic and kinetic parameters for all of the possible species and reactions involved. As an indication of the size and complexity of the problem, a model has been presented for the growth of GaAs from TMGa and arsine which involves 24 chemical species and 45 reactions (20). In spite of this complexity, a more limited set of thermodynamic and kinetic data may be used to predict trends that can be used to improve the control of the MOCVD process. Thermodynamic values are used to determine the equilibrium state for a given set of conditions and the driving force for a particular reaction. Thermodynamics can be used to determine the concentrations of reactants at equilibrium, to predict which reactants will yield improved results, establish trends for changes in process parameters, and predict maximum equilibrium growth rates. Kinetics describe the rate at which a reaction approaches equilibrium. Because MOCVD is not an equilibrium process, the actual growth rates and compositions of the resulting deposits are influenced by the kinetics, ie, reaction rates, mass transport, and the importance of individual reaction steps, of the system. For the MOCVD of GaAs, a complete thermodynamic and kinetic description is very complex (20). Most evidence indicates that surface adsorption of the reactants takes place first and then surface reactions occur to form the semiconductor material under normal MOCVD growth conditions (17,20).

Transport Phenomena. One significant though often overlooked kinetic aspect of MOCVD is the influence that gas flow or hydrodynamics has on mass transport and uniformity of deposition. Because of the complexity of these effects, they are not intuitive and are often overlooked in the design of an MOCVD reactor. Both simplified and complete numerical descriptions of the hydrodynamics of certain types of MOCVD reactors have been published (7,16,21–23). These articles suggest reasons for using certain types of reactors as well as design criteria that can be used to minimize the effects of temperature gradients, buoyancy driven circulation, and reactor boundaries. The three types of reactors that are commonly used in MOCVD are pictured in Figure 3. The reactor types are commonly referred to as the vertical reactor, the planetary rotation reactor, and the horizontal reactor. The overall gas flow direction is indicated in each reactor. Another type of reactor geometry used in production is the barrel reactor which can hold a large number of substrates. It is configured with a vertical flow, and the substrates are placed on the sides of a rotating substrate holder (16).

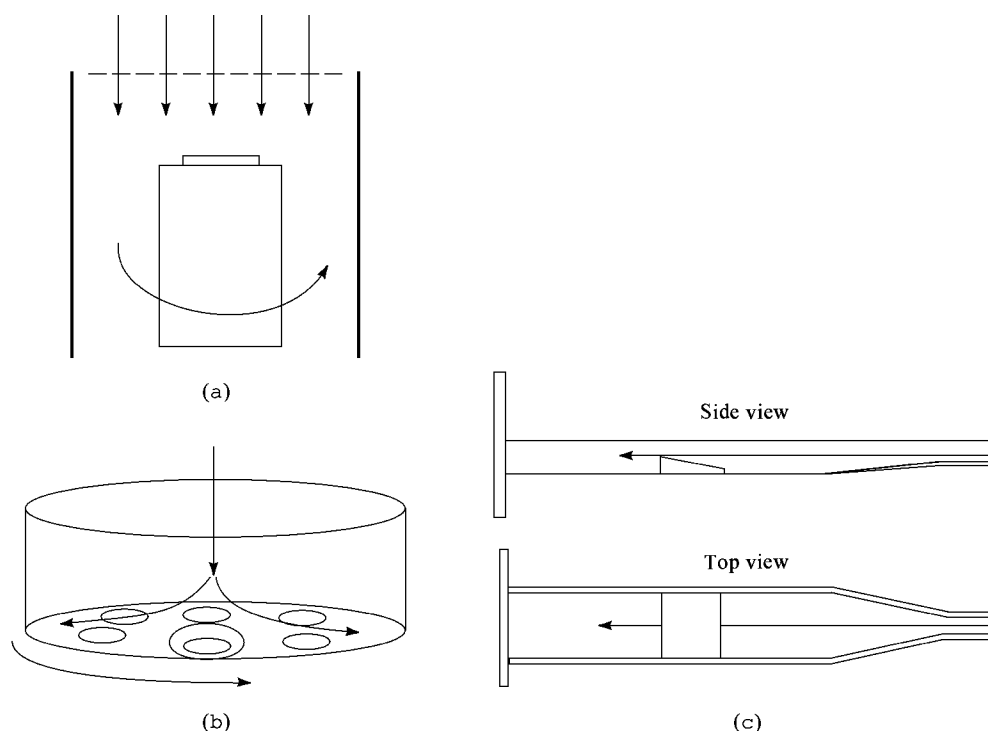


Fig. 3. Schematic of three commonly used types of MOCVD reactors where the arrows indicate gas flow: (a) vertical rotating disk where (— — —) represents an inlet to promote a laterally uniform gas flow, (b) planetary rotation, and (c) horizontal.

In Situ Diagnostics. Some of the complications that complex fluid flow causes in MOCVD can be overcome through the use of *in situ* diagnostic instrumentation. The use of *in situ* diagnostics is primarily aimed at reproducibility in growth rates and composition. A number of techniques have been used as *in situ* monitors in MOCVD. These include x-ray diffraction using synchrotron radiation (24), infrared and ultraviolet spectroscopy (25,26), ellipsometry and reflection difference spectroscopy (27), spectral reflectance (28), and ultrasonics (29). Some of these techniques are primarily for research and can be used to gain further insight into the reaction mechanisms of MOCVD; others such as ultraviolet spectroscopy, spectral reflectance, and ultrasonics have been implemented into real time process control.

Materials. As understanding of the MOCVD process has increased, it has become possible through the synthesis of new chemicals to improve the uniformity and purity of materials grown by MOCVD and to increase the growth parameter range over which device quality materials can be grown (30). Originally the sources consisted of the trimethyl or triethyl Group III compounds and Group V hydrides. Because of the increased demand for improved materials and increased understanding, new chemicals have been designed which take into consideration the decomposition mechanisms of the sources and how this impacts on the incorporation of impurities, particularly carbon, into the epitaxial layer. The replacement of methyl Group III sources with ethyl groups has led to a marked decrease in carbon incorporation in AlGaAs grown by MOCVD (31). A further reduction of carbon can be achieved through the use of trimethylaminealane which has no carbon bound directly to Al (32). Tertiary butyl arsine and phosphine are replacements for Group V hydrides which are highly toxic (33). Their use has resulted in the growth of high quality materials. In this case, the addition of carbon to the Group V atom has not resulted in increased carbon incorporation. This is explained by the decomposition mechanism which results in the formation of AsH_2 . The AsH_2 reacts just like arsine on the growth surface and does not result in higher carbon incorporation (34). The use of new organometallic sources has increased researchers' process design parameters and will enable them to expand the growth capabilities of MOCVD.

Reaction Mechanisms. An integral part of the increased understanding of the MOCVD process has come about from a number of investigations into the reaction mechanisms for MOCVD. In a series of experiments (7) done to clarify the gas-phase decomposition of the Group III organometallics, it was found that the organometallics probably decompose predominantly on the surface through an interaction with other surface species. The existence of methyl and Group III radicals in the vapor phase has been demonstrated by infrared and ultraviolet spectroscopies (25,26). Temperature programmed desorption indicates the presence of mono- and dimethylgallium on the surface of GaAs at temperatures below 400°C . Evidence for the existence of methylene groups on the surface explains why only a small amount of the absorbed methyl groups incorporate as carbon into the growing epitaxial layer (35). X-ray diffraction patterns of the surface of GaAs during growth indicate that the surface reconstruction present during MOCVD is not the same structure as that found in MBE at high vacuum (36). This evidence can be explained using a reaction mechanism for MOCVD that involves the separate absorption of trimethylgallium and arsine onto the GaAs surface and their decomposition through a series of surface reactions (7,16). At high temperatures, the gas-phase decomposition of the reactants is an important side reaction.

The above reaction mechanism can be used to explain the observed growth rate behavior of GaAs (7,16). The growth rates for GaAs from trimethylgallium and arsine have three different temperature regimes (7,16,17). The low growth rates encountered in the low ($<500^\circ\text{C}$) temperature regime are explained by a lower decomposition rate of trimethylgallium on the GaAs surface. At an intermediate ($500\text{--}800^\circ\text{C}$) temperature, the growth rate is independent of temperature and limited by the mass transport to the surface. At higher ($>800^\circ\text{C}$) temperatures, the growth rate decreases with increasing temperature and reflects an increased rate of desorption of the reactants from the surface before they can interact to form GaAs (7,16). The effect of surface orientation and arsine pressure on the growth rate is explained by the surface decomposition mechanism. Different surface sites that exist on different crystallographic orientations result in different decomposition rates (37).

The existence of methyl and methylene radicals on the surface during MOCVD growth explains the incorporation of C as an impurity. Other impurities that are typically found in MOCVD materials include O, Si, Ge, and Zn. The primary source of these impurities is the starting reactants. The incorporation of these impurities is dependent on growth temperature and V:III ratio. Attempts to improve the quality of the material include the use of alternative sources as previously discussed and chemical purification of the sources. In spite of these impurities, high quality GaAs and InP have been grown by MOCVD with net carrier concentrations $<10^{13}\text{ cm}^{-3}$ and mobilities as high as 210,000 and 300,000 cm^2/Vs , respectively.

Dopants. Carbon has also been found to be an effective dopant for the III:Vs grown by MOCVD because of its relatively high solubility in Al-containing materials and its low diffusivity (7). Other dopant sources used in MOCVD include methyl or ethyl compounds Se, Te, Be, Zn, and Cd and hydrides S and Se. Tetraethyltin and silane or disilane are also used as Group IV dopant sources. An unusual dopant source is bis-cyclopentadienyl magnesium. The Group VI sources are *n*-type dopants for the III:Vs, whereas the Group 2 and 12 (II) elements are *p*-type dopants. Si and Sn are usually *n*-type; carbon is usually *p*-type. However, because these elements can occupy either lattice site under some conditions they are amphoteric and can be self-compensating. The carrier concentration of the resulting doped layer is proportional to the partial pressure of the dopant source molecule. The variation of carrier concentration with temperature, growth rate, and V:III ratio depends on the incorporation mechanism for the individual dopant. Hydrogen has also been shown to compensate some dopants (38). It can usually be removed by annealing at high temperature.

Heterostructures and Superlattices. Although useful devices can be made from binary compound semiconductors, such as GaAs, InP, or InSb, the explosive interest in techniques such as MOCVD and MBE came about from their growth of ternary or quaternary alloy heterostructures and superlattices. For the successful growth of alloys and heterostructures the composition and interfaces must be accurately controlled. The composition of alloys can be predicted from thermodynamics if the flow in the reactor is optimized. Otherwise, composition and growth rate variations are observed

across the surface of the substrate. As discussed, the surface reactions can be complex and lead to deviations from the predicted compositions. The temperature profiles and hydrodynamics of the reactor also affect the resulting composition from the effects of the transport of the reactants to the surface and the incomplete decomposition of the reactants. These effects are well illustrated by the GaAsP system where the amount of P incorporated varies dramatically with temperature due to the large difference in decomposition temperatures of arsine and phosphine (7,39).

Phase Separation. There are two common complications that occur when preparing ternary alloys of the III–V semiconductors. The first is the existence of a two-phase region in the phase diagram for the alloy of interest instead of a continuous solid solution. This results in phase separation or spinodal decomposition of the alloy. Phase separation occurs when the bond energy is very dissimilar between two of the constituents so that the entropy of mixing is not large enough to overcome the difference in the enthalpies of formation of the binaries and their solid solutions. This is expressed by a positive free energy of formation for the reaction of the two binaries to yield the desired ternary.

Ordering in Ternary Alloys. The second complication is the existence of ordering in ternary alloys. Ordering, or the formation of a nonzincblende crystal structure, has been found to occur in most of the III–V alloys grown by MOCVD. The most common form of ordering in these materials results in a CuPt-like phase where a superlattice with a translational symmetry of double the normal lattice occurs in the $\langle 111 \rangle$ direction. Ordering has a pronounced effect on the optical properties of alloys. Usually the higher the degree of ordering, the lower the band gap. This can be advantageous in some instances where a lower band gap might be desired, such as for long wavelength detectors in the InAsSb system, but deleterious in others such as AlGaInP where shorter wavelengths are desired. Ordering is not observed for alloys grown by liquid-phase epitaxy (LPE) and probably results from the existence of reconstructions of the surface during growth which results in the preferential incorporation of different elements at different surface sites (7,40).

Abrupt Interfaces. The successful use of heterostructures in a device requires that they be prepared with abrupt interfaces. The growth of an abrupt interface can be accomplished only by an abrupt change in the gas-phase composition. This can be accomplished by a growth interruption, the use of high flow rates, low growth rates, gas switching, and various combinations of these techniques. Through the use of gas switching sequences, one can prepare interfaces with a desired chemical composition. This is illustrated by the growth of InGaAsP on InP where the interface can consist of InAs, GaAs, GaP, or various ternary bonds. These different interfaces result in different strain profiles at the interface and different electrical and optical properties for very thin layers (41,42).

The growth of thin superlattices and heterostructures with very abrupt layers and interfaces is probably the most significant development in MOCVD device technology. This ability has resulted in vastly improved device performance for lasers, high speed transistors, and detectors. This improvement resulted from the ability of MOCVD and MBE to grow heterostructures with low interfacial recombination velocities, which are directly related to the ability to prepare interfaces with very low defect densities that are nearly atomically abrupt (7). A transmission electron micrograph of a superlattice of alternating layers of InAsSb and InSb is shown in Figure 4. This high magnification micrograph indicates the sharpness of the interfaces that can be obtained with MOCVD. Monolayer abrupt interfaces have been claimed for GaAs–AlGaAs (7).

MOCVD is a very versatile growth technique with the capability of growing most of the III–V and II–VI semiconductors. High quality heterostructure devices using atomically abrupt layers have been prepared by MOCVD and are being produced commercially. The exploration of new material growth, growth mechanisms, uniformity, material purity, and large-area deposition will continue as the commercial applications of MOCVD expand.

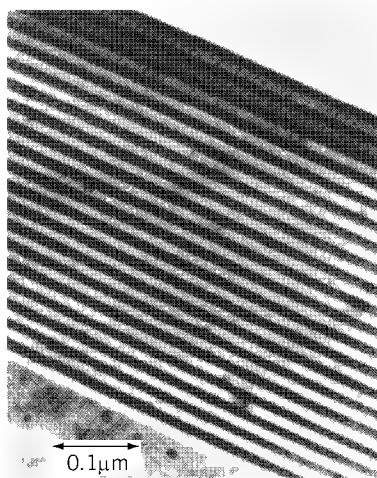


Fig. 4. A transmission electron photomicrograph of an InAsSb–InSb superlattice grown by MOCVD.

Electronic Devices

The main advantages that compound semiconductor electronic devices hold over their silicon counterparts lie in the properties of electron transport, excellent heterojunction capabilities, and semi-insulating substrates, which can help minimize parasitic capacitances that can negatively impact device performance. The ability to integrate materials with different band gaps and electronic properties by epitaxy has made it possible to develop advanced devices in compound semiconductors. The hole transport in compound semiconductors is poorer and more similar to silicon. For this reason the majority of products and research has been in *n*-type or electron-based devices.

Some of the original applications of compound semiconductors were in the development of diodes for microwave power and oscillator applications, including impact ionization avalanche transit times (IMPATTs) and transferred electron devices (TEDs). More recently, development efforts have focused on three terminal devices including heterojunction bipolar transistors (HBTs) and field-effect transistors (FETs). Presently, HBTs and FETs are the devices of choice for analogue, microwave, and digital circuit design in compound semiconductors. Much more recently, a newer class of devices called quantum effect devices (QEDs) has been the subject of intense research. QEDs include resonant tunneling diodes (RTDs) and single-electron transistors (SETs). There is a great deal of speculation surrounding the functionality of these devices as present device geometries shrink to the $<0.1 \mu\text{m}$ range, where quantum effects are expected to dominate device behavior.

DIODES: IMPATTs AND TEDS

Both IMPATTs and TEDs have been popular devices for use in microwave and millimeter wave power applications. These devices have been increasingly replaced by HBTs or FETs. Both devices are generally homostructures, using a single compound semiconductor, typically GaAs or InP (43). However, the doping profiles often make use of epitaxy, requiring sharp profile with low background and excellent control of low doping levels ($\sim 10^{15} \text{ cm}^{-3}$). Descriptions of IMPATTs and TEDs are available (43–45).

The IMPATT diode relies on the use of a nonuniform doping profile to tailor the electric field of a device under bias. A high electric field region is used to cause a local avalanche multiplication by impact ionization. A large drift region is then used to control the propagation of the carrier pulse. Figure 5a shows the doping and electric field profiles of an IMPATT device or Read diode (46). A d-c bias is applied to generate an electric field just below the critical field for impact ionization. A smaller a-c bias causes the impact ionization to occur on the positive swing; on the negative swing the current is

collected. This results in a 180° phase lag between the peak current and the positive peak in the a-c voltage.

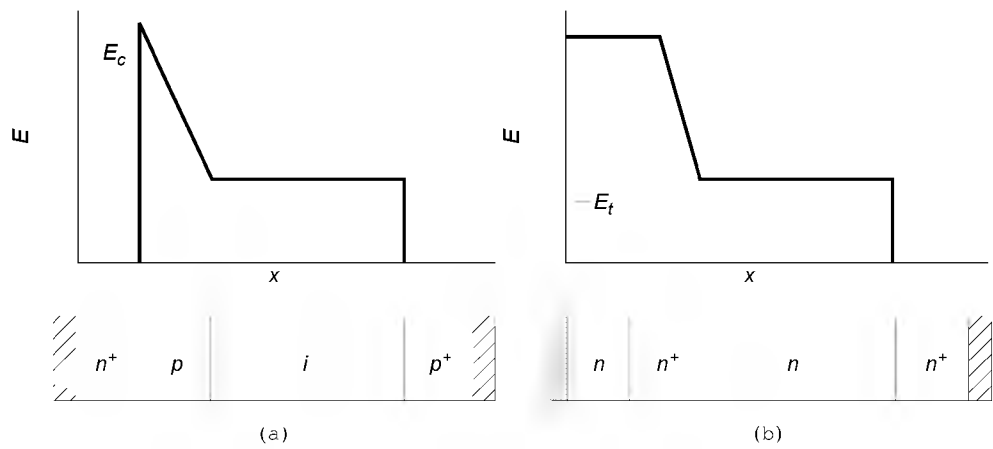


Fig. 5. Structures and electric fields for (a) Read IMPATT diode and (b) Schottky cathode TED where ▨ represents ohmic and ▩ Schottky contact, and *i* denotes the intrinsic or undoped region of the semiconductors.

TEDs rely on the negative differential velocity inherent in the compound semiconductors with a second energy minimum that has a lower saturated velocity. Semiconductors that exhibit this behavior include GaAs and InP. An example velocity field curve is shown in Figure 2. The TED is biased so that the electric field is above the point where the maximum velocity occurs, E_p or within the negative differential velocity regime. Figure 5b shows the doping and electric field profiles for a Schottky barrier contacted TED (43). Random charge fluctuations in the device result in a local space charge density that can propagate across the transit region of the device. A complete description of this phenomenon is available (45). As with the IMPATT, a combination of d-c bias and an a-c signal is used to initiate and propagate the charge pulse.

Although IMPATTs and TEDs are effective in producing microwave power at frequencies between 10 and 300 GHz, they are presently being replaced by FETs and HBTs. The reasons for this are twofold. Both IMPATTs and TEDs are extremely noisy when compared to FETs and HBTs, which can be a problem for many applications. The second issue revolves around integration. FETs and HBTs are being integrated into complete microwave circuits on a chip or millimeter wave-integrated circuits (MMICs). The advantages of this fabrication technology make FETs and HBTs the more attractive alternatives when possible.

FIELD-EFFECT TRANSISTORS

The basic operating principle of a FET is the modulation of a gate electrode bias controlling the current between two other electrodes, the source and drain, with minimal gate current. In silicon microelectronics the metal oxide semiconductor FET (MOSFET) dominates the electronic device community. The gate or controlling electrode is separated from the silicon conducting channel by a thin oxide layer. This prevents any gate leakage current for all but the most severe operating conditions. In compound semiconductors, it has proven very difficult to fabricate a MOSFET-like transistor or a metal insulator semiconductor FET (MISFET). In a MOSFET the insulator is silicon dioxide, which can form a well-behaved interface with silicon. The fabrication of a well-behaved or trap-free interface between an insulator and a compound semiconductor is difficult and has hampered efforts to produce MISFETs. Subsequently, compound semiconductor FETs rely on direct metal to semiconductor gates, and are much more likely to experience gate leakage at nominal operating conditions.

There are many ways to fabricate FETs in semiconductors with each separate technique employing a new acronym. Device types are divided as to how the electrons are confined in a channel, whether a semi-insulating barrier exists between the channel and the gate, and how the device is fabricated, ie, either self-aligned (SA) or nonself-aligned (NSA). One simple electrical distinction is whether the FET is conducting at zero volts on the gate. If a conducting channel exists between the source and drain when the gate voltage is zero volts, the device is said to be depletion mode or d-mode, requiring a negative gate voltage to turn the device off. If the device is off with zero volts on the gate, then the device is called enhancement mode or e-mode, requiring a positive bias on the gate to turn the device on.

Fabrication of FETs. Figure 6 shows the two basic geometries for compound semiconductor FETs, recessed gate and self-aligned. There are some similarities between the two structures. In both cases it is desirable to make the source and drain ohmic contacts as low resistance as possible. Additionally, the material between the contacts and the gate represents a parasitic resistance which needs to be minimized or the device performance suffers. To accomplish this the semiconductor is highly doped to achieve a low resistance. In a self-aligned FET (Fig. 6b) this doping is accomplished by ion implantation after the gate metal is in place. Because a high temperature anneal is required to activate the implant, the gate metal is usually a refractory metal such as tungsten. After implant activation the ohmic contacts are deposited and annealed. In the recess gate geometry (Fig. 6a) the semiconductor is made conducting by implantation anneal or epitaxy before the metal contacts are in place. The source and drain contacts are then deposited and annealed to provide a low resistance contact. Before deposition of the gate, a light patterned wet etch is used to thin the conducting layer. The same photoresist that is used to mask the wet etch is then used to define the gate metal lift-off. In this way, the gate is aligned to the recess, but is not self-aligned to the whole structure. Although the gate is shown midway between the source and drain contacts, varying the gate placement can have beneficial effects on such properties as breakdown and drain-source output conductance (47).

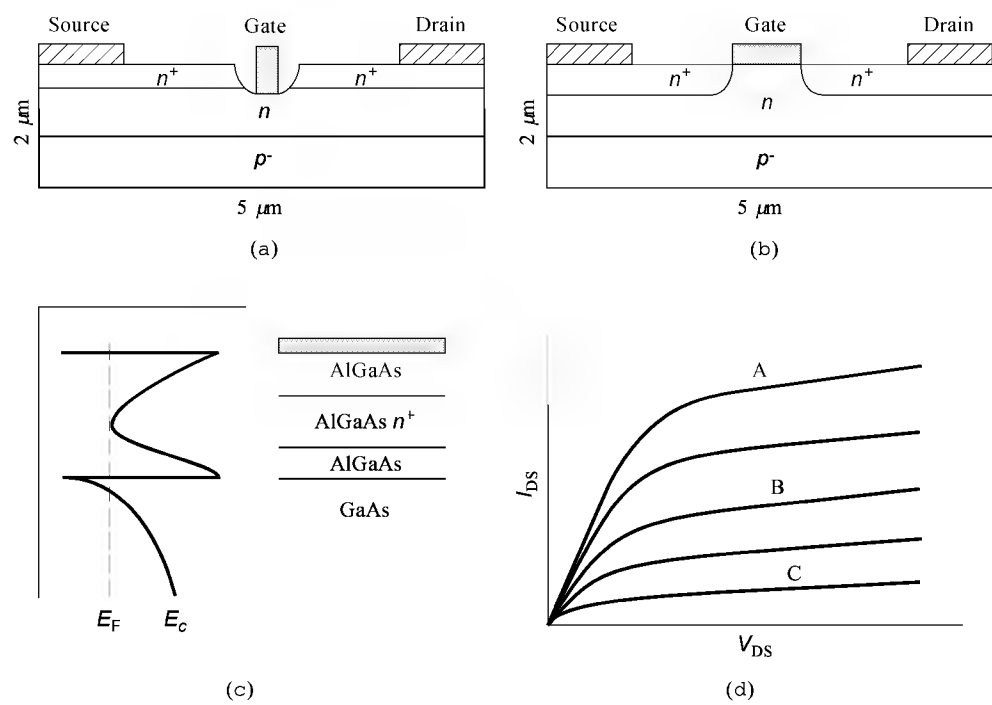


Fig. 6. (a) Recess gate FET geometry and (b) self-aligned FET geometry where $\square$ represents ohmic and $\blacksquare$ Schottky contact; (c) layer structure and conduction band profile for a typical heterojunction FET (HFET), where E_c is conduction band edge energy and E_F is Fermi energy; and (d) typical I–V characteristics for a d-mode FET where I_{DS} is the source-to-drain current, V_{DS} is source-to-drain voltage, and gate voltage, V_g for A is >0 ; B, $V_g = 0$; and C, $V_g < 0$.

The channel region of the device is located directly below the gate metal. Except for fringing effects, it is the charge in the channel region that is modulated by a voltage applied to the gate. The semiconductor material beneath the gate can be homogeneous as in a metal semiconductor FET (MESFET) or junction FET (JFET), or it can be a heterojunction structure grown by epitaxy. In a MESFET the gate is placed directly on the channel material, with the doping defined either by implantation or epitaxy. In a JFET a p – n diode is used as the gate, with the gate contact acting as a ohmic contact to the p -region, in the case of an n -channel device.

Epitaxial devices usually enable shallower or sharper dopant profiles than are possible with implantation. This can be extremely important in very short gate length devices. Epitaxy is also used to exploit the properties of heterojunctions in a FET structure. Figure 6c shows the epitaxial structure for a typical HFET (heterojunction FET). There are many different ways of arranging the layers within a HFET. The two main approaches have tried to exploit modulation doping or semi-insulating layers. In a modulation doped structure a wide band gap material is doped while a narrow band gap material is left nominally undoped. The dopants are placed within a few tens of nanometers (nm) of the interface. The electrons transfer to the lower band gap material, where high electron density can be achieved with spatial separation from the parent donor atoms. This allows the material to have high electron mobility which can improve device performance (48). Typical heterojunctions used include AlGaAs–GaAs and InAlAs–InGaAs. These devices are often called high electron mobility transistors (HEMTs) or modulation doped FETs (MODFETs). The narrow band gap layer can be semi-infinite (see Fig. 6c) or a quantum well, with more wide band gap material below. Both approaches have advantages and disadvantages (49). In a different approach, the heterojunction is used to reduce gate leakage. If wide band gap material is left undoped, it acts as an insulator in a MIS-like structure. However, because the wide band gap material does have a finite conductivity and band gap, often small improvements in gate leakage are all that can be expected. In these cases the channel doping is located within the narrow band gap material. These devices are often called heterojunction insulated gate FETs (HIGFETs) or doped channel FETs (DCFETs).

FET Device Characteristics. The current between the source and drain of a FET is the product of the electron density in the channel (below the gate) and the velocity of the electrons. The gate modulates this current by controlling the electron density in the channel. The velocity of the carriers can be controlled by increasing the bias on the drain, with the source typically grounded. When a voltage is applied to the gate the amount of charge on the gate is changed. In response to this, the charge in the channel redistributes. For n -type devices a negative gate bias causes a depletion of the channel electrons, whereas a positive gate bias causes an accumulation of channel electrons. The situation is reversed for a p -type device with holes in the channel. A JFET works in the same way, with a p – n diode acting as the gate instead of a metal–semiconductor Schottky contact. The electrons in the device that are not under the gate are not modulated by the gate bias and populate connection paths between the channel and the source and drain ohmic contacts.

Although the relationship between the channel charge and gate bias can be complicated, it can be simply approximated as a capacitor structure. In this simplification the following equation holds, where q is the electron charge, n_{ch}

$$qn_{ch} = C_{GC}V_{GS}$$

is the electron density in the channel, C_{GC} is the gate–channel capacitance, and V_{GS} is the gate–source voltage. Typically, this approximation is valid over the mid-range of operating conditions. At very low gate biases the device is said to be near threshold, the point where the device current turns on, and this relationship is no longer valid. At very high gate biases, the channel charge may saturate and the gate begins to leak. High performance devices try to modulate the most charge with the smallest changes in gate bias. This requires a high gate–channel capacitance, which is achieved by placing the gate close to the conducting channel.

The electron velocity within the channel is primarily controlled by the drain-to-source bias and the gate length. For very long devices (gate lengths $>2\text{ }\mu\text{m}$), the electron velocity is equal to the product of the electric field and the mobility. In the long gate length regime the source-to-drain current can be expressed as follows, where $\beta = q\mu\epsilon W/2L_Gd_{GC}$, L_G is the gate length, d_{GC} is

$$I_{DS} = \beta(V_{GS} - V_T)^2$$

the gate-to-channel distance, μ is the electron mobility, ϵ is the dielectric constant, W is the device width, and V_T is the device threshold which is dependent on the details of the doping (45). For short gate length devices ($<1\text{ }\mu\text{m}$), the electric field due to the drain-to-source bias is very high and the electrons tend to travel at their saturated velocity, v_{sat} . When the electrons can be assumed to be traveling at their saturated velocity, the current is approximately the product of v_{sat} and qn_{ch} . Typical d-c I–V characteristics are shown in Figure 6d.

A popular metric of the speed performance of a FET is f_p , the maximum frequency of unity current gain (49). For short gate length devices operating at the saturated velocity this metric can be expressed as follows, where $g_m = \partial I_{DS}/\partial V_G$

$$f_t \approx \frac{g_m}{2\pi C_G} \approx \frac{v_{\text{sat}}}{L_G}$$

is the device transconductance, and C_G is the total gate capacitance. This implies that for high speed performance short gate lengths and high saturated velocities are desirable. Extremely short gate length ($L_G < 0.1\ \mu\text{m}$) devices have exhibited f_t values in excess of 300 GHz (50). More typical performance for 0.5- μm gate length devices is approximately 30–60 GHz (49). This is the driving reason behind the constant decrease in FET gate lengths. Additionally, for digital applications the power consumption drops for shorter gate lengths. However, as the devices become shorter the scaling properties become very important. In order to maintain good modulation control of the channel, the channel doping needs to become shallower as the gate length is reduced. The ratio of L_G to d_{GC} needs to be maintained above four for good scaling behavior (49). Additionally, very short gate length devices have low breakdown voltages due to the very high electric fields present. This generally necessitates a trade-off between speed and power performance.

Because f_t is also proportional to the saturated velocity, significant effort has been made in developing materials with increased electron saturated velocity. This has included InGaAs on InP substrates and even InAs devices. However, the higher saturated velocity generally comes with a lower band gap which also implies easier breakdown. In addition to the concept of saturated velocity, there has been work in the area of ballistic transport. The device gate lengths are so short that the electrons do not have time to reach steady-state transport, and it is predicted that the electron velocity can greatly exceed the saturation values for short distances (45).

HETEROJUNCTION BIPOLAR TRANSISTORS

The FET and the heterojunction bipolar transistor (HBT) are the dominant electronic device technologies in compound semiconductors. Whereas the FET is operated as a voltage controlled current amplifier, the HBT is used as a current controlled current amplifier. Although the operating principles are quite different the output characteristics are similar in that a small control signal is used to switch a larger magnitude output signal on and off. The gain of the device is an attractive feature for both analog and digital applications. There has been a great deal of discussion around the issue of whether HBTs or FETs are more suited to particular applications. However, it is sufficient to say that both devices are extremely useful and have application in a wide range of circuits.

In silicon technology the HBT can be fabricated from Si–SiGe heterojunctions, but the homojunction Si bipolar junction transistor (BJT) is more common. The BJT and the HBT are nearly identical devices, with the HBT incorporating at least one heterojunction interface between device regions. Homojunction BJTs in compound semiconductors are not an area of active interest, as the HBT offers large performance gains necessary for high speed and high power applications.

In an HBT the charge carriers from an emitter layer are transported across a thin base layer and collected by a third layer called the collector. A small base current is present which includes the carriers that did not successfully cross the base layer from the emitter to the collector. The FET is a unipolar device making use of a single charge carrier in each device, either electrons or holes. The HBT is a bipolar device, using both electrons and holes in each device. The emitter and collector layers are doped the same polarity (n - or p -type), with the base being the opposite polarity (p - or n -type). An HBT with a n -type emitter is referred to as a n - p - n device; a p - n - p device has a p -type emitter. The n - p - n transistors are typically faster and have been the focus of more research. For the sake of simplicity, the following discussion will focus on n - p - n transistors.

HBT Design. HBTs are similar to FETs in that poor fabrication can result in large parasitic resistance and capacitances which can dominate the device performance. As such, there are many different fabrication technologies for HBTs, making a complete overview impossible herein. Therefore, this discussion will focus on the more common aspects of HBT fabrication. A typical emitter-up HBT is shown in Figure 7a. The layers are deposited by epitaxial techniques. In this example the collector is deposited first, followed by the base, and then the emitter. There has also been work in collector-up structures in an attempt to minimize excess base–collector capacitance (51). In this example the collector and base are of the same narrow band gap material, with a wide band gap emitter. Typical material combinations are AlGaAs–GaAs, InP–InGaAs, and InAlAs–InGaAs. This structure is referred to as a SH-HBT or a single heterojunction HBT. Devices have been fabricated which use a wide band gap material for the collector and are referred to as DH-HBTs or dual heterojunction HBTs. The use of a wide band gap material in the collector can increase the device breakdown voltage and therefore increase the maximum available power from the device.

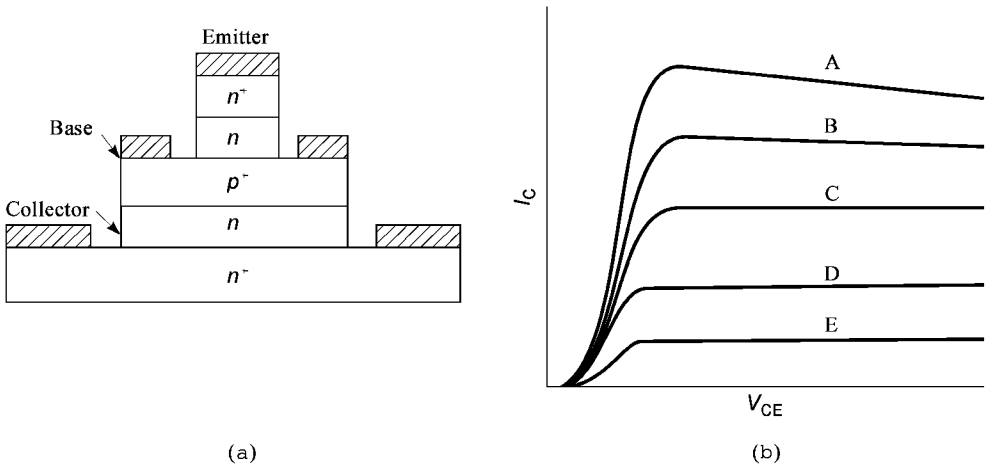


Fig. 7. (a) Generic structure of an emitter-up HBT and (b) typical I–V characteristics of an HBT, where A shows I_{B1} ; B, I_{B2} ; C, I_{B3} ; D, I_{B4} ; and E, I_{B5} .

The doping in both the emitter and collector are typically nonuniform. Near the region where an ohmic contact is made, the doping is raised to reduce any parasitic resistances. However, closer to the emitter–base or collector–base junction the doping is modified to provide the proper I–V characteristics of the individual n - p diodes. The base is typically thin ($<100\ \text{nm}$) and is very heavily doped ($>10^{19}\ \text{cm}^{-3}$). This is to facilitate ohmic contact formation and lower the sheet resistance of the base layer. Most of the device action occurs directly under the emitter contact, and the base layer between the emitter contact and base contact acts as a parasitic resistance. The emitter layer is defined using either wet or dry etching, often with selective etches to ensure stopping in the correct place. The exact details are particular to the device technology, self-aligned vs nonself-aligned. Very high doping or self-aligned contacts are used to minimize this resistance. The very high doping of the base layer can often introduce unwanted dopant diffusion (52). This must be accounted for in the fabrication process, often by leaving a thin ($<10\ \text{nm}$) undoped layer between the base and emitter to allow for dopant diffusion (53). Because the emitter contact is typically very small, its contact resistance is of paramount performance. In addition to the metallization schemes discussed elsewhere, it is not unusual to see the emitter layer graded down to a very narrow band gap material to facilitate ohmic contact formation (54).

The active region of an HBT is directly beneath the emitter contact, where the injected electrons cross the base and enter the collector. Because the emitter is etched away in regions to allow base contacting, the emitter–base junction is typically of minimum size. However, the same is not true for the base–collector junction. In addition to the region directly below the emitter, the entire length of the base out to the base ohmic contact forms a capacitor with the collector. This parasitic capacitance has a large negative effect on device performance. A technique for minimizing this capacitance is the use of ion

implantation (qv) to isolate the collector region (55). By proper choice of ion and energy it is possible to render the collector region near the base insulating, without significantly affecting the base resistance. This dramatically reduces the parasitic base–collector capacitance.

HBT Device Characteristics. The HBT consists of two back-to-back *n*–*p* diodes. In the most typical configuration the emitter–base diode is forward biased, with the collector–base diode reverse biased. Because the current in a forward-biased *n*–*p* diode is exponentially dependent on the bias, small changes in the emitter–base voltage result in large changes in the emitter current. The current across the emitter–base junction is a combination of the electrons injected into the base and the holes injected into the emitter. If the diode was semi-infinite to each side, the electron current density, *J_n*, could be expressed as follows (44), where *q* is the electron charge, *V* is the bias across the diode, *kT*

$$J_n = \frac{qD_n n_{po}}{L_n} \left(\exp \left(\frac{qV}{kT} \right) - 1 \right)$$

is the thermal energy, *D_n* is the electron diffusion coefficient in the base, *L_n* is the electron debye length in the base, and *n_{po}* is the equilibrium electron concentration in the *p*-type base material. The key aspects of this equation are (1) the current is exponential with bias and (2) the current is proportional to *n_{po}*. As the best performance results if all of the emitter–base voltage is dropped across the junction and the base layer is typically very thin, a heavy doped base layer is desirable to minimize the base resistance away from the junction. However, a large *p*-type doping in the base lowers *n_{po}* because *np = n_i²* (44), where *p* is the hole density and *n_i* is the intrinsic carrier concentration, which is temperature dependent. A high *p*-type doping density in the base, therefore, would result in a lower electron current and a higher hole current, exactly the reverse situation desired. However, if a heterojunction *n*–*p* diode is used, an additional barrier to the hole current comes from the valence band discontinuity (56). In this case the ratio of the electron-to-hole current is as follows (56), where

$$\frac{I_n}{I_p} \approx \frac{N_E}{P_B} \exp \left(\frac{\Delta E_v}{kT} \right)$$

N_E is the emitter *n*-type doping, *P_B* is the base *p*-type doping, and Δ *E_v* is the valence band discontinuity. Valence band discontinuities are usually in the range of 0.1–0.3 eV. Because the exponential can be very large, this allows the ratio of the emitter-to-base doping to vary considerably. This is helpful in allowing the base doping to be increased thereby lowering the base resistance. In addition to the hole current injected into the emitter, some of the emitter electron current suffers from recombination within the base layer which adds to the base current. Typically this is a small fraction of the total current. A key parameter for HBTs is the current gain or β, which is the ratio of the emitter current to the base current. This can be a very large number in HBTs as it is proportional to the exponential of the valence band discontinuity. Typically, however, β is not maximized, as the overall structure takes advantage of the exponential to vary doping levels to reduce parasitic effects. Figure 7b shows an example of the d-c characteristics for a common emitter HBT.

An approximation for the maximum frequency of unity current gain, *f_β* in an HBT can be expressed as follows (53), where τ_E is the emitter charging

$$\frac{1}{2\pi f_t} = \tau_E + \tau_{BT} + \tau_{CT} + \tau_C$$

time, τ_{BT} is the base transit time, τ_{CT} is the collector transit time, and τ_C is the collector charging time. The emitter charging time, τ_E, can be expressed as the RC product of the emitter resistance and the emitter–base capacitance. The base transit time, τ_{BT}, is equal to the width of the base divided by the carrier velocity. The collector transit time, τ_{CT}, is equal to the width of the collector depletion region divided by the carrier velocity. The collector charging time, τ_C, can be expressed as the RC product of the base–collector capacitance and the collector series resistance. Although these equations are simplified, they do provide insight as to why the series resistances and excess capacitances are important to minimize. State-of-the-art HBTs have exhibited *f_t* > 100 GHz.

QUANTUM-EFFECT DEVICES

As modern devices become smaller in pursuit of higher speed and lower power operation, they are entering length scales where quantum effects are expected to become important, if not dominant. Instead of trying to avoid these quantum effects in conventional devices, an alternative approach is to exploit these effects in developing new devices (57–59). Although conventional devices often improve with shrinking size, small dimensions are a requirement for the successful operation of quantum-effect devices (QEDs). This field is in its infancy with many issues to be resolved including how best to incorporate quantum effects into devices that can be used in a complete system. Because the behavior of a QED may be significantly different from conventional devices, work is also being carried out in system architecture issues, as the best way to utilize these unconventional devices is unclear (60). Many QED concepts have increased functionality over the basic on–off behavior of FETs or HBTs. This could allow the design of complicated systems with decreased numbers of devices thereby increasing speed and further reducing power consumption.

Electron tunneling is the basis for the resonant tunneling diode (RTD), a two-terminal device exhibiting on–off–on behavior. The RTD uses epitaxy to achieve the small dimensions necessary to control the tunneling, whereas the lateral dimensions can be >1 μm, allowing conventional lithography to be used for fabrication. Other QEDs which use a horizontal geometry, relying on ultrasmall lithography, include single-electron transistors and electron interference devices. Single-electron transistors (SETs) are devices in which the addition or subtraction of a single electron is sufficient to control the operation. There have been a number of other device technologies proposed in compound semiconductors that have yet to be proven experimentally. These include electron interference devices (61) and a novel technology for quantum cellular automata (QCA) (62).

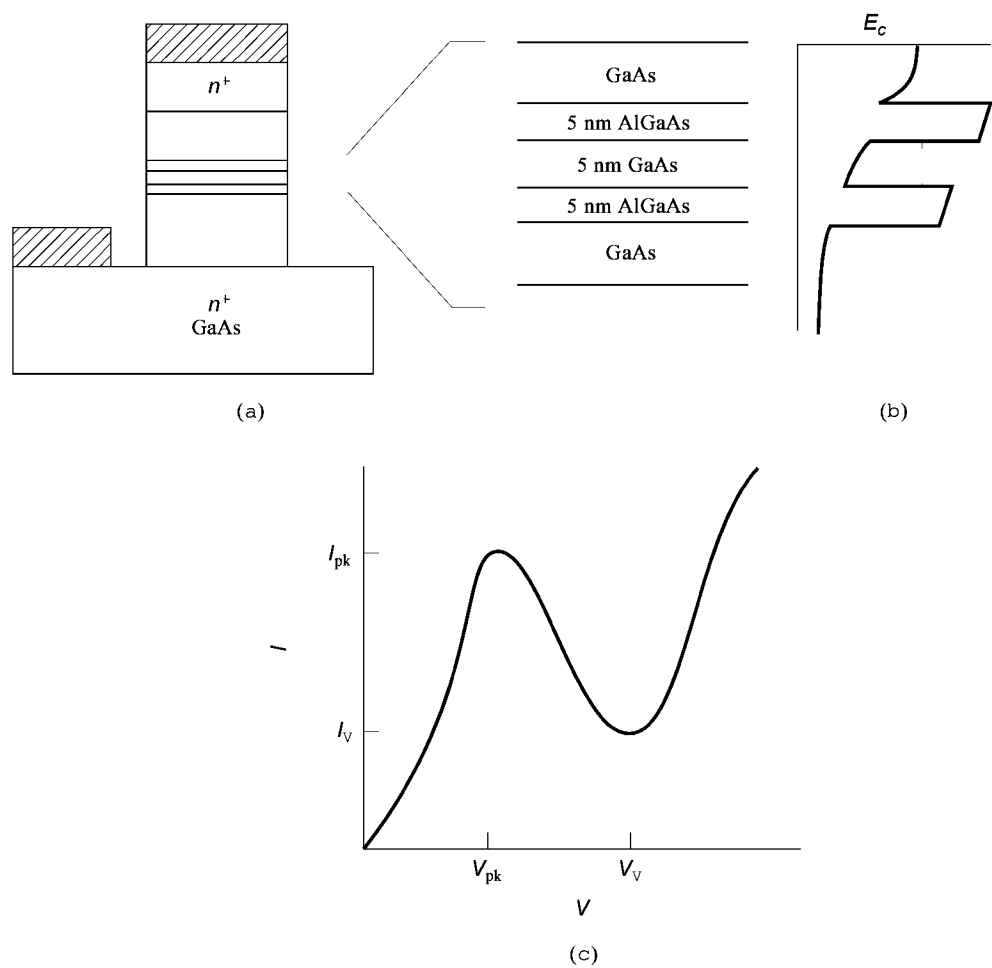


Fig. 8. (a) Structure of a typical resonant tunneling diode (RTD); (b) conduction band diagram for the barrier structure where (—) represents the resonant state; and (c) typical I–V characteristics for a RTD.

RESONANT TUNNELING DIODES

The structure of a AlGaAs–GaAs resonant tunneling diode (RTD) is shown in Figure 8a. Typical heterojunctions for RTD fabrication include AlGaAs–GaAs, InAlAs–InGaAs, and InAs–AlSb. The key component of the RTD is the thin double-barrier structure grown by epitaxial techniques. Typical length scales are <5 nm. Figure 8b shows the conduction band along the vertical direction through the double-barrier structure. The two AlGaAs wide band gap barriers form a resonance state in the confined GaAs layer. Electron transmission through the double-barrier structure is limited to those electrons that can conserve both energy and momentum as they tunnel from the emitter, through the barrier, and into the collector (63). Typical I–V characteristics are shown in Figure 8c. Initially the electrons in the emitter do not have enough energy to tunnel to the collector and the current is low. At a bias of V_{pk} a large number of emitter electrons satisfy the tunneling constraints and can tunnel through the barrier, resulting in a high current. As the bias is further raised, the emitter and barrier fall out of resonance and the tunneling current reaches a minimum. With further bias increases thermionic emission over the barrier dominates. The key characteristics of the RTD are the peak current, I_{pk} , the voltage at peak current, V_{pk} , the minimum or valley current, I_v , and the voltage at the valley current, V_v . Investigations into InAlAs–InGaAs and Sb-based RTDs are based on the improvements on peak current and the peak-to-valley ratio (PVR) (64). The best room temperature PVRs of 11 have been achieved for InAs–AlSb (64). In general the improvements come from higher conduction band discontinuities between the barrier material and the confined well material. Lower temperatures also improve the sharpness of the tunneling characteristics due to the reduced thermal spreading of the electron distribution in the emitter.

RTDs have exhibited microwave oscillations at frequencies in excess of 700 GHz (65). Unfortunately, the available power output is very low. Although circuit design with two terminal devices is more difficult than with three terminal devices, RTDs are being explored for use in digital logic (66), multivalued logic (67), memory circuits (68), and analog circuits (69). Work is also in progress to develop three-terminal resonant tunneling transistors (RTTs) (49,61).

SINGLE-ELECTRON TRANSISTORS

The greatest potential application for single-electron devices lies in digital circuits. However, a number of other applications exist, including current standards and ultrasensitive electrometers (70,71). SETs are not unique to compound semiconductors, and in fact a great deal of work has been carried out in other material systems, including Al–AlO_x–Al tunnel junctions. A review of single-electron phenomena is available (72).

The basic operating principle of a SET lies in how much energy is required to add or remove a single electron from the system. Figure 9a gives the structure for an AlGaAs–GaAs SET. These devices are also referred to as Coulomb blockade devices. The AlGaAs–GaAs heterojunction is used to fabricate a vertically confined two-dimensional electron gas. Gates A, B, C, and D are then biased negatively to isolate region E from the surrounding electrons. Region E is typically very small (<100 nm²) and is occupied by a small number of electrons (~1–1000s). The depletion regions formed between gates A–C and B–C act as the inlet and outlet for electrons traveling through region E. By lowering the negative potential on gates A or B, single electrons can be allowed to enter or leave region E. Modulating gate D allows the energy of the system (region E), to be modified. The energy to add or remove a single electron is $e^2/2C$, where e is the electronic charge and C is the total capacitance of the charge island (region E). If the island's electron occupancy is an integer, n , then there is a block to charge moving from region F to G. However, if the occupancy is $n + 1/2$, oscillating between n and $(n + 1)$, then it is possible for electrons to travel from region F to G. Additionally, because the electron occupation of region E is changed by the voltage on gate D, the conductance is periodic in gate voltage as shown in Figure 9b.

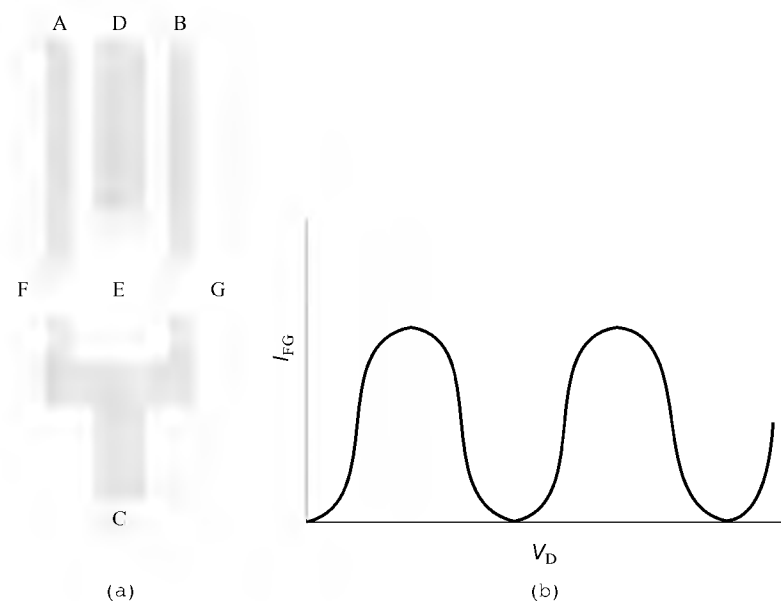


Fig. 9. (a) Top view of a Coulomb blockade device (single-electron transistor) patterned on an AlGaAs–GaAs heterojunction, where  the surface gate; (b) I–V characteristics for a Coulomb blockade device.

Due to thermal effects such devices must operate at temperatures well below the electron charging energy of $e^2/2C$. With state-of-the-art fabrication technology, the capacitance is typically of the order 10^{-16} F, which requires temperatures below 1 K. Even with further miniaturization, it is unlikely that these devices will be feasible at room temperature. Even so, there has been work in modeling this type of device for use in digital circuits (73).

There is another class of SETs that do not make use of the energy to add single electrons to a system. These devices use the trapping of individual electrons to modify the conductance of a more conventional FET channel (74). By trapping charge in polysilicon grain boundaries, these devices have been operated at room temperature with very promising electrical characteristics. This device operates on the electron's ability to electrostatically deplete the channel, not on the energy required to trap the electron.

Photonic Devices

A technology for which compound semiconductors are uniquely suited is that of photonics, which describes devices that generate, amplify, detect, propagate, transmit, or modulate light. This field therefore covers a wide range of systems, including light-emitting diodes (LEDs), lasers and optical amplifiers, detectors, waveguides such as optical fibers, lenses, and other optical components, and optical modulators. Although silicon has contributed much to photonics technology with its applicability to visible and near-infrared (ir) light detection systems, the compound semiconductors, with their wide range of primarily direct bandgaps, enable the realization of efficient light-emitting devices, namely laser diodes and LEDs, and extension of detector technology into the far ir region of the spectrum.

The relevance of photonics technology is best measured by its omnipresence. Semiconductor lasers, for example, are found in compact disk players, CD-ROM drives, and bar code scanners, as well as in data communication systems such as telephone systems. Compound semiconductor-based LEDs utilized in multicolor displays, automobile indicators, and most recently in traffic lights represent an even bigger market, with approximately \$1 billion in annual sales. The trend to faster and smaller systems with lower power requirements and lower loss has led toward the development of optical communication and computing systems and thus rapid technological advancement in photonics systems is expected for the future. In this section, compound semiconductor photonics technology is reviewed with a focus on three primary photonic devices: LEDs, laser diodes, and detectors. Overviews of other important compound semiconductor-based photonic devices can be found in References 75–78.

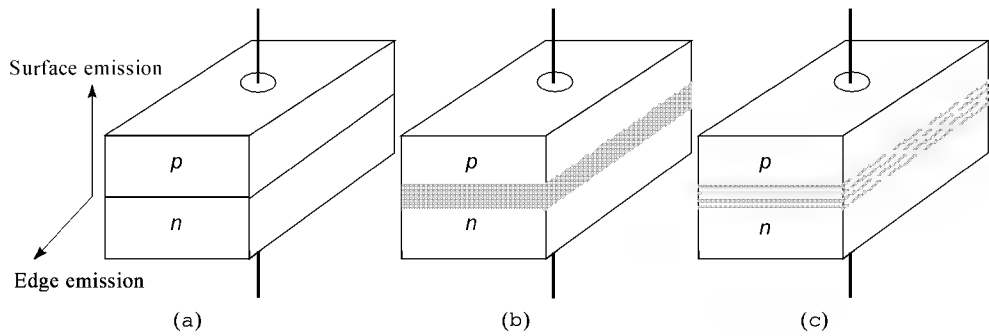


Fig. 10. Schematic of various LED and laser diode structures where  signifies material of a lower energy band gap: (a) homojunction, (b) double-heterojunction (DH), and (c) multiquantum well (MQW) structures.

LEDs

A light-emitting diode (LED) is a forward-biased *p–n* junction in which the applied bias enables the recombination of electrons and holes at the junction, resulting in the emission of photons. This type of light emission resulting from the injection of charged carriers is referred to as electroluminescence. A direct band gap semiconductor is optimal for efficient light emission and thus the majority of the compound semiconductors are potential candidates for efficient LEDs.

The most simple LEDs are homojunction devices, composed of the same bulk material with *n*- and *p*-type doped regions, as shown schematically in Figure 10a. Greatly increased performance has been achieved with the development of heterojunction devices which combine two or more semiconductor materials of different band gaps. As shown in Figure 10b, the different band gaps of the heterostructure enable the formation of potential wells in the junction region to better confine the electrons and holes which results in more efficient photon generation. Advances in semiconductor growth technologies such as MOCVD and MBE allow for the growth of more complex structures, such as the multiquantum well structure, shown in Figure 10c. In this structure, the center well of the double heterostructure LED is replaced with one or more very thin (<10 nm) layers which result in further improved carrier confinement as well as changes in the energy and wavelength of emission.

In addition to the structural variations, there are also two principal LED geometries, namely the surface-emitting and edge-emitting geometries shown in Figure 10. For the surface-emitting LED, light is emitted in a direction perpendicular to the junction of the device. In most cases this is the top surface of the device, although if the structure is grown on a substrate material that does not absorb the emitted photons, emission can be obtained from both the top and bottom surfaces of the device. In contrast, the edge-emitting LED emits light in a direction parallel to the device junction which, for a cubic crystal geometry, is usually cleaved $\langle 110 \rangle$ facets for a structure grown on a $\langle 100 \rangle$ substrate. This geometry is similar to that used for edge-emitting laser diodes and is generally less efficient than the surface-emitting geometry, and has an asymmetric beam profile. One clear advantage of the edge-emitting geometry is that it enables the development of superluminescent diodes which have greater output power and broader spectral output than surface-emitting LEDs.

Whether a particular LED will be useful in photonics applications depends critically on its wavelength, spectral bandwidth, external quantum efficiency, and output power. The emission wavelength and spectral bandwidth are governed primarily by the band gap of the semiconductor materials that make up the LED, with larger band gap materials emitting at shorter wavelengths, because the band gap is inversely proportional to the wavelength, and generally narrower spectral linewidths. The external quantum efficiency of the device is equal to the ratio of the produced photon flux to the injected electron flux and depends on both how efficient the material is at generating photons (internal quantum efficiency) as well as how efficiently the photons can travel from the junction and out of the material without being absorbed or reflected back (transmission efficiency). The total output power, P , of the LED is determined by the following relation where I is the injected current in the device, c is the speed of

$$P_0 = \frac{\eta_{ex} h c I}{\lambda e}$$

light, e is the charge of the electron, h is Planck's constant, and η_{ex} is the previously described external quantum efficiency. Thus, for an ideal LED, the output power is proportional to the injected current.

The fact that compound semiconductors have a wide range of band gaps has enabled the development of LEDs from the ultraviolet (uv) to the mid-ir region of the spectrum. Starting at the shorter wavelengths, GaN is a Group III–V semiconductor material which is suitable for uv and blue-based LEDs (79). This new technology has overwhelmed the performance of commercial SiC-based blue LEDs with an order of magnitude improvement in output efficiencies (approaching 4%) and estimated device lifetimes as high as 100,000 hours at an emission wavelength of 450 nm. Group II–VI materials based on ZnSe have also been proven to enable efficient LEDs in the blue and green regions of the spectrum, with significantly narrower spectral bandwidths than the GaN LEDs and slightly lower efficiencies (80,81). To date (ca 1996), the II–VI LEDs have suffered from shorter lifetimes than the GaN materials, with projected lifetimes of approximately 1000 hours. One of the most important applications for these blue LEDs is that of multicolor displays. The GaN LED is in use as the critical third component of red–blue–green (RBG) multicolor displays, the other wavelengths having been developed previously out of other material systems.

LED technology in the rest of the visible region of the spectrum is dominated by GaP, GaAsP, AlGaAs, and AlGaInP materials. Although both GaP and GaAsP are indirect gap materials, they are often doped with impurities to enhance radiative efficiency, typical examples being nitrogen-doped GaP (GaP:N), zinc and oxygen-doped GaP (GaP:Zn,O), and nitrogen-doped GaAsP (GaAsP:N). These GaP and GaAsP LEDs span the deep red (700 nm for GaP:Zn,O) to the green (555 nm for GaP). Although not extremely efficient (~0.7% external quantum efficiency for 630 nm red GaAsP:N, 0.1% for 555 nm green GaP), these materials compose up to 90% of the LED market because they can be grown with the mature and cost-effective technique of liquid-phase epitaxy (LPE) (82). AlGaAs-based LEDs are appropriate for long wavelength visible (650–700 nm) LEDs, and are used in traffic lights and car tail-lights. External quantum efficiencies of >16% have been achieved at 650 nm. Direct gap AlGaInP materials, which cover the red to green regions of the spectrum, are promising for shorter wavelength visible LED applications (82,83). Significant advances in AlGaInP-based LEDs have been made by bonding the material to a transparent GaP substrate to enable light emission from both sides of the device (84). This technology has resulted in LED performance that exceeds all other commercial technologies for the green to red spectral regime.

The visible LEDs dominate the LED market, but longer wavelength LEDs also have applications. Near-ir LEDs (780–1000 nm) are used in short haul communication systems and are typically made of GaAs, AlGaAs, and InGaAs materials. The 1.1–1.6-μm region of the spectrum is also relevant for optical communication systems and LEDs emitting in this region are based on InGaAsP materials. Finally, InAsSb and related alloy structures have been developed into LEDs covering the mid-ir (2–5 μm) region of the spectrum. These wavelengths are useful for gas monitoring systems, such as those used to monitor pollution.

Beyond these commonly used double-heterostructure (DH) surface-emitting LEDs, there are several more advanced LED systems. Resonant cavity LEDs (RCLEDs) are structures where the light-emitting (active) region is grown in between two semiconductor-based mirror stacks, thus forming a resonant optical cavity (85,86). Typically the cavity is highly asymmetric, with one mirror having a reflectivity close to 100% while the other has a reflectivity of typically < 97%. The effect of the cavity is to restrict the spectral bandwidth of the light emission to one of the modes of the cavity and also to enhance the light emission from the LED. Thus, RCLEDs can have higher spectral purity and higher intensities than standard LEDs and may be used in applications such as short-distance fiber optic communication systems, where narrower spectra translate to less dispersion in the optical fiber. Particular examples include visible (670 nm) AlGaInP-based RCLEDs (87) and InGaAs-based 940-nm RCLEDs (88), both designed for a spectral width of approximately 5 nm.

Other applications such as wavelength division multiplexing (WDM) and some spectroscopic applications require high powers and a broader spectral width than typical LEDs. In this case, a superluminescent LED is used, which typically has an edge-emitting geometry with cleaved facets and is essentially the same structure as a laser diode. The difference is that some loss mechanism, usually in the form of a tilted mirror or an absorbing region along the length of the LED, is introduced that prohibits the device from actually lasing. This allows a relatively high level of current injection without lasing, such that much higher powers and broader spectral outputs are achieved than from standard LEDs. Superluminescent LEDs have been reported throughout the near-ir region of the spectrum (89–91) and are sold commercially. At the communication wavelengths, as an example, 80- and 140-nm linewidths have been achieved at 1.3 and 1.5 μm, respectively (90).

LASERS

First developed in 1962, the semiconductor laser diode has rapidly matured to become the dominant laser source used in photonics technologies, primarily due to its small size, low power requirements, manufacturability, and robustness (92–94). Just as for LEDs, the wide array of primarily direct band gap compound semiconductors has enabled semiconductor laser development from the blue to the mid-ir region of the spectrum. The semiconductor laser diode is similar to the LED in that it utilizes the injection of charged carriers to emit light. In addition, the laser diode employs an optical cavity formed by mirrors, often just the cleaved facets of the semiconductor, to allow for optical feedback. The optical feedback enables stimulated emission to occur, in which reflected photons stimulate the emission of additional photons of the same energy and phase. Injecting current into the device provides gain to the system which is the mechanism that allows for the multiplication or amplification of photons. The photons that travel in the cavity also experience a certain loss, due to absorption in the cavity or transmission through the mirrors. When the gain of the laser system is greater than the loss of the system, lasing occurs.

Because lasers are governed by stimulated emission, their properties are distinct from those of LEDs which emit spontaneous emission. In particular, the fact that stimulated emission generates photons of equal phase is responsible for lasers having coherence. This is important for applications such as holography and enables improved focusability of light. As the mirrors of the cavity determine a preferred direction for photon propagation, the laser emission is in the form of a well-defined beam, as opposed to the isotropic emission from LEDs. In contrast to the relatively broad spectral output of the LED, the laser has one or more sharp emission lines, arising from optical modes that propagate due to the boundary conditions imposed by the laser cavity. Finally, the fact that the losses of the cavity must be overcome before lasing can occur results in a well-defined threshold injection current below which lasing cannot occur.

Several heterostructure geometries have been developed since the 1970s to optimize laser performance. Initial homojunction lasers were advanced by the use of heterostructures, specifically the double-heterostructure device where two materials are used. The ability of the materials growth technology to precisely control layer thickness and uniformity has resulted in the development of multi-quantum well lasers in which the active layer of the laser consists of one or more thin layers to allow for improved electron and hole confinement as well as optical field confinement.

Like the LED, the laser diode has both edge-emitting and surface-emitting geometries, with edge-emitting lasers being by far the most common. For the edge-emitting laser, light propagates in the plane of the wafer and the mirrors are typically fabricated by cleaving the wafer along a crystallographic plane. Although the reflectivity of these cleaved mirrors is only about 30%, it is sufficient to support lasing in broad area lasers that have high gain. To reduce mirror loss, the facets can be coated with a dielectric multilayer stack to enhance reflectivity of the mirrors. The cavity length of edge-emitting lasers is usually on the order of 250–1000 μm.

Another type of edge-emitting laser commonly used in communications applications is the distributed feedback (DFB) laser. In this device, the cleaved mirrors that are commonly used are replaced by corrugations fabricated in the semiconductor material, as shown schematically in Figure 11. These corrugations result in a periodic refractive index variation that enables coherent feedback of the light propagating in the cavity (95). This complicated mirror has definite advantages over a cleaved mirror in that it only allows one of the many optical modes of the laser cavity to propagate, resulting in single-mode (therefore single-frequency) output. As the optical fibers used in communication systems have significant wavelength dispersion at 1.55 μm, the sharp spectral profile of the DFB laser results in less distortion of the optical signal.

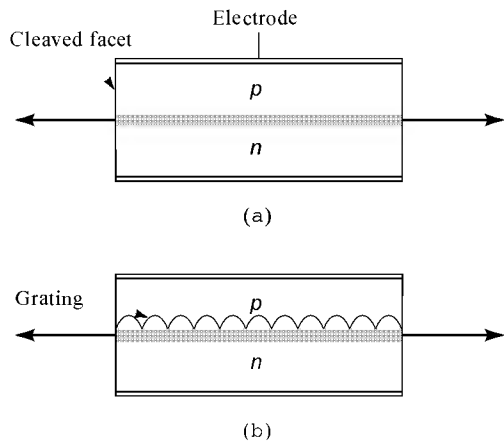


Fig. 11. Schematic of edge-emitting laser diodes where the arrows represent the direction of laser emission and ■ represents the active region: (a) standard structure with cleaved facets for mirrors and (b) distributed feedback (DFB) laser that employs coherent reflection from a grating to generate optical feedback.

The surface-emitting laser geometry has two basic designs. One is an in-plane laser similar to that of an edge-emitting laser, but that uses some form of reflection to direct the light output perpendicular to the plane of the laser cavity. An example of this is shown in Figure 12a where etched 45° mirrors are fabricated at the output facet of an edge-emitting laser to direct the light. This type of laser is often referred to as a planar cavity surface-emitting laser (PCSEL). The other design, a surface-emitting design that is distinctly different from the edge emitter, is the vertical cavity surface-emitting laser (VCSEL) for which both the cavity and light propagation is perpendicular to the plane of the wafer (Fig. 12b) (96). The cavity is typically designed to be λ/n in thickness, where λ is the laser wavelength and n is the material index of refraction at that wavelength. This translates to about 200 nm (2×10^{-7} m) for a visible emitting VCSEL. Because the cavity is so thin, the photons do not experience much gain as they propagate through the cavity. Therefore, the losses in the system, particularly the mirrors, must be very small.

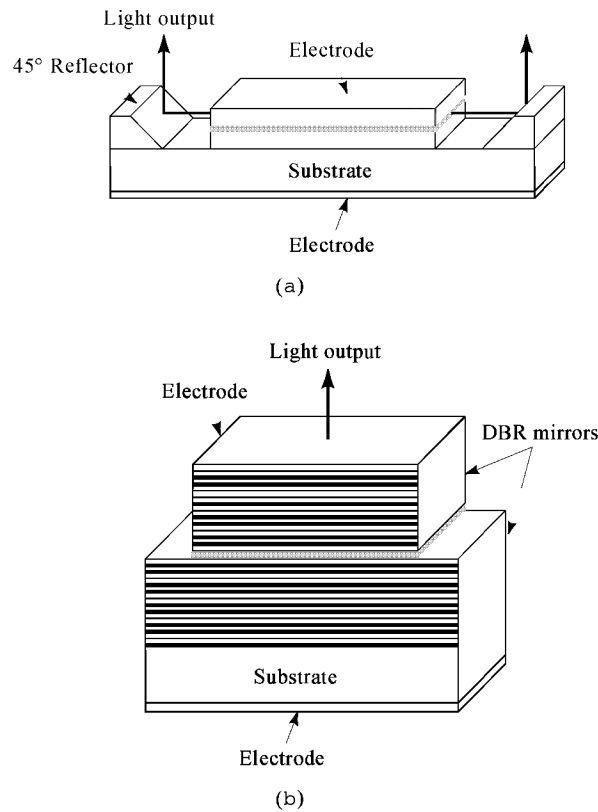


Fig. 12. Schematic of surface-emitting laser diodes where ■ represents the active region: (a) planar cavity surface-emitting laser diode (PCSEL) with 45° etched reflectors and (b) vertical cavity surface-emitting laser diode (VCSEL) with semiconductor-based multilayer mirror stacks grown into the structure. DBR = distributed Bragg reflector.

To achieve low loss, multilayer semiconductor mirrors are grown on each side of the laser cavity. Each layer is of thickness $\lambda/4n$ to form a coherent reflection of light at a designed wavelength, and as many as 50 layers are needed to achieve reflectivities on the order of 99.99%. These mirrors are called distributed Bragg reflectors (DBRs) and are one of the critical components of the VCSEL. Although VCSELs typically are not as high power as edge-emitting lasers, they have several advantages due to their vertical geometry. These include a circular, low divergence beam that is optimal for coupling to optical fibers, amenability to dense two-dimensional array fabrication due to their geometry, and the ease of wafer level testing because cleaving is not necessary to create mirrors.

Compound semiconductor-based laser diodes operate in the blue–green, red, and near- to mid-ir region of the spectrum. One of the areas of most recent developments is that of blue–green laser diodes. Interest in these short wavelength lasers arises primarily from optical data storage applications where the shorter wavelengths translate to the ability to read or write more information on a disk. In early 1996, Nichia Chemical Industries reported the first demonstration of a room temperature GaN-based laser diode, with emission at 415.6 nm in the blue region of the spectrum (97). The GaN-based materials field is rapidly advancing and improvements in GaN laser diode performance as well as expansion into the uv and green regions of the spectrum may be seen in the near future.

Another material system for which blue and green laser diodes have been demonstrated is that of ZnSe and related Groups II–VI materials (3,98). These II–VI materials are primarily grown by molecular beam epitaxy (MBE). The first demonstration of electrically injected blue–green edge-emitting lasers, observed at low temperatures and under pulsed conditions, occurred in 1991 (99). Rapid progress has extended the technology to room temperature continuous wave (CW) emission in the blue–green region of the spectrum (510–520 nm) (100–102), with pulsed operation down to approximately 463 nm (103). One problem with these II–VI-based laser diodes is that they suffer irreversible degradation with CW lifetimes on the order of one hour (104). ZnSe-based VCSELs have been demonstrated with optical pumping (105,106).

As of early 1996, commercial laser diodes were restricted to emission in the red region of the spectrum. The dominant material for red laser diodes is AlGaInP, which has demonstrated room temperature pulsed lasing in broad area lasers down to emission wavelengths approaching 610 nm (107), with longer wavelengths (670–690 nm) giving significantly better performance. AlGaInP high power laser diode arrays have been developed with up to 12 W emitted at 640 nm (108). Red VCSELs were first achieved from AlGaInP materials in 1993 (109), with more recent performance at 690 nm of up to 8 mW output power (110). Applications for red lasers include laser printers, bar-code scanners, optical data storage systems, and plastic fiber-based communications systems.

GaAs, AlGaAs, and InGaAs materials are used in laser diodes with emission from 750 to 1000 nm. One of the principal technologies in this wavelength region is that of compact disk players and CD-ROM drives which use 780-nm AlGaAs DH lasers (92). Although these lasers have become an industry standard, other materials have been developed to generate lasing in this wavelength region. One such material is the quaternary InGaAsP grown on GaAs substrates, which can be grown to cover the same band gap range (1.42–1.92 eV or 870–650 nm) as the AlGaAs materials (111). Near-ir semiconductors are also employed to pump erbium-doped fiber amplifiers that are used in long distance communication systems. InGaAs lasers are used to supply the 980-nm laser emission wavelength, which is optimal for absorption into the doped fiber. Although the wavelengths of these 780–980-nm lasers are not optimized for fiber-based communication systems, they are routinely used as a low cost solution for short distance communication applications as well.

Near-ir (750–1000 nm) VCSELs based on GaAs, AlGaAs, and InGaAs materials are the most mature of the VCSEL technologies, and dramatic improvements in device performance have been achieved just through 1995. At present, InGaAs-based 980-nm VCSELs have been demonstrated with greater than 50% electrical to optical power conversion efficiency (112) and threshold currents less than 100 μ A (113,114). Output powers as high as 23 mW have been obtained (115). Advances in GaAs-based 850-nm VCSELs include CW output power of 59 mW from a 40- μ m diameter device and lasing up to 200°C (116).

The most important laser wavelengths for communication applications is the 1.1–1.7 μ m region in the near-infrared. This arises from the fact that glass optical fibers used in data communication systems have minimum dispersion at 1.3 μ m and minimum loss at 1.55 μ m (see FIBER OPTICS). The primary material system for laser diodes in this wavelength region is InGaAsP grown on InP substrates, where $\text{In}_{0.73}\text{Ga}_{0.27}\text{As}_{0.58}\text{P}_{0.42}$ and $\text{In}_{0.58}\text{Ga}_{0.42}\text{As}_{0.90}\text{P}_{0.10}$ alloys are used to achieve emission at 1.3 and 1.55 μ m, respectively. AlGaInAs alloys on InP substrates have also been used to achieve lasing in this region (117). VCSELs have been demonstrated both at 1.3 (118) and 1.55 μ m (119) since 1994, although the room temperature output powers remain low.

Laser sources that emit in the mid-ir region of the spectrum (2–5 μ m) are useful for detection of trace gases because many molecules have strong absorption bands in that region. Other applications include remote sensing and laser radar. Semiconductor lead–salt (IV–VI) lasers that operate CW at a temperature of 200 K and emission wavelength of 4 μ m are commercially available; however, they have relatively low output powers (<1 mW) (120). Group III–V quaternary materials have also been developed as mid-ir laser diodes, with the strained quantum well GaInAsSb–AlGaAsSb laser demonstrating up to 1.3 W CW output power at 2 μ m (121). Many other ternary and quaternary III–V structures have been developed to enable longer wavelength emission. InAsSb–AlInAsSb strained quantum well structures are one example, with CW lasing at 3.9 μ m achieved up to an operating temperature of 128 K (122).

DETECTORS

Whereas there are several types of photon detectors, the most common compound semiconductor-based devices are photoconductors and photovoltaic devices (75,76,123). With the photoconductive detector, a change in the resistivity of the semiconductor material as light is absorbed is measured. The spectral range of detection is limited on the long wavelength end by the semiconductor band gap (intrinsic detectors) or the binding energy of an impurity in the semiconductor (extrinsic detectors) and the current–voltage characteristics are nominally linear. The photovoltaic devices, often called junction devices or photodiodes, employ a *p–n* junction. The long wavelength limit of detection is determined by the band gap of the semiconductor material for most types of detectors. The intersubband detector is an exception. The absorption of light results in the generation of electrons and holes which are subsequently swept out of the depletion region of the junction. This carrier transport acts to forward bias the junction and to produce either an open-circuit voltage or a short-circuit current which is measured. The current–voltage characteristic is not linear but rather shows rectifying behavior (123). Many of the semiconductor materials that are used in detectors, such as InSb, InAs, HgCdTe, PbTe, PbSe, and PbTe, have been developed as both photoconductive and photovoltaic detectors.

The most critical parameters that govern the performance of the detector are quantum efficiency, responsivity, and response time. The quantum efficiency η is the probability that a photon incident upon the detector will generate an electron–hole pair which contributes to a detectable current. Factors such as the reflection of the photons from the surface of the detector, generation and recombination of the electron–hole pair outside of the junction, and absorption efficiency of the material reduce the quantum efficiency from the theoretical limit of 1. The responsivity *R* is a related parameter that relates the incident photon flux to the electric current and is measured in W^{−1} A. Finally, the response time of the detector is simply the time it takes for the current to be generated and depends on factors such as the resistance and capacitance of the photodiode and drive circuitry, and the mobility of carriers in the material. As detectors are often called on to operate in high speed applications, the response time of the detector is a critical parameter. Many applications, including high speed data communication systems, require the detector to respond to the intensity variation of light that is modulated up to a frequency of >10 GHz (10¹⁰ Hz).

Silicon-based detection systems are the dominant technology for single-element and array detectors in the near-uv to near-ir spectral region (350–1000 nm), with a quantum efficiency at least 20% greater than that of compound semiconductors (approaching a value of 1 in the 700–800 nm range). The advantage of compound semiconductors is that they can cover a much wider spectral range than silicon, operating from the uv to the far-ir region of the spectrum. Applications for long wavelength ir (LWIR) detectors that operate in the 8–12 μ m region are numerous and include systems for navigation, weather, and pollution monitoring, remote sensing, night vision, and navigation. The most commonly used material system is HgCdTe–CdTe that

operates in the 3–17- μm region of the infrared and is utilized in commercial imaging arrays (124). Mid-ir detection is accomplished by InSb arrays, and near-ir detector technologies include InGaAs–InP materials that operate in the 1.3–1.6- μm region (124,125). Additional near-ir materials are quaternary materials such as InGaAsP–InP and AlGaAsSb–GaSb which operate in the 0.9–1.7- μm region.

Developments have also been made on GaAs–AlGaAs quantum well infrared photodetectors (QWIPs) (126). GaAs has a large band gap with respect to InSb and HgCdTe, and does not absorb infrared radiation longer than approximately 900 nm by direct band gap absorption. However, a GaAs multiquantum well structure can be designed that has confined energy levels only a few meV apart (94). This allows for the resonant absorption of infrared radiation equal in energy to the energy separation between the excited energy state and the ground state of each of the quantum wells, as shown schematically in Figure 13. Because the energy spacing can be controlled by the thickness of the quantum wells, the detection wavelength region can be tuned to relatively long wavelengths. As an example, InGaAs–GaAs QWIPs have been developed with detectivity out to 19 μm (127). Although HgCdTe-based detectors still have higher detectivities at longer wavelengths ($\lambda > 8\ \mu\text{m}$) and at operating temperatures in the 4–77 K region, the GaAs-based QWIPs show higher detectivity at temperatures below 40 K (128).

Many of these material systems have been developed into avalanche photodiodes (APDs) which involve not only the generation of carriers with the absorption of light but also the amplification of the generated carriers (75,76). This is accomplished by applying a very high reverse bias to the detector, resulting in a high enough internal electric field inside the material to generate additional carriers by impact ionization. These detectors therefore have a higher responsivity than more conventional photodiodes, but suffer from increased noise.

As for LEDs, substantial improvement in detector performance can be achieved by placing the LED structure inside a resonant cavity. In particular, the resonant cavity photodetector can achieve high quantum efficiencies in a relatively narrow spectral region (86). In one study, a mirror was designed to allow for resonant absorption of two wavelengths: 730 and 910 nm (129). This type of multiple wavelength detector shows promise for wavelength division multiplexing (WDM) systems where wavelength multiplexing enables enhanced transmission capacity.

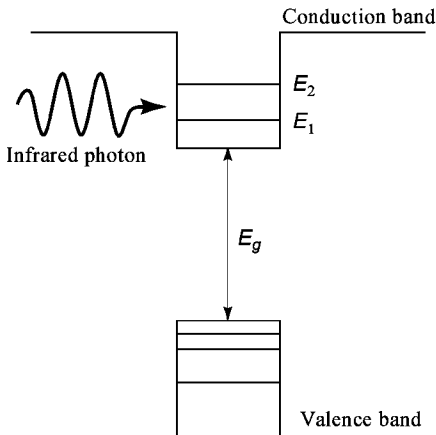


Fig. 13. Absorption between confined energy levels in a quantum well infrared photodetector (QWIP). The energy difference (E_2-E_1) between the confined energy levels in a quantum well may be designed such that it is resonant with ir radiation. The band gap, E_g is much greater, therefore direct band gap absorption does not occur.

Device Fabrication Technology

WET ETCHING

Compound semiconductor processing makes extensive use of etching for, eg, active area definition, gate recess etching, and waveguide formation. Wet etching can provide a clean, damage-free surface with good control of both etch depth and lateral undercut. Most wet etches are isotropic which may limit their usefulness in high aspect ratio submicrometer applications where straight wall profiles are required. With the proper combination of semiconductors and etchants, excellent selectivity can be achieved to extend the utility of wet etches to the nanometer range.

Table 4. Wet Etches for Compound Semiconductors

Semiconductor	Etch	Typical rates, $\mu\text{m min}$	Reference
GaAs ^a	H ₃ PO ₄ :H ₂ O ₂ :H ₂ O	0.01–4.0	137
	C H ₈ O ₇ :H ₂ O ₂	0–0.3	133
	H ₂ SO ₄ :H ₂ O ₂ :H ₂ O	3–14	137,138
AlGaAs	NH ₄ OH:H ₂ O ₂	0.1	139
	C H ₈ O ₇ :H ₂ O ₂	0–0.2	133
InGaAs	C H ₈ O ₇ :H ₂ O ₂	0–0.14	133
InP	HCl	6	140
	HCl:H ₃ PO ₄	0.1–1.0	141
	Br–CH ₃ OH	0.4	140
InAs	C H ₈ O ₇ :H ₂ O ₂	0.09	142
InSb	I ₂ :CH ₃ OH		143
GaSb	HNO ₃ :HF:CH ₃ COOH		143
CdTe	C H ₈ O ₇ :H ₂ O ₂	0.9	142
	Br:CH ₂ OHCH ₂ OH	0.08	144
ZnTe	HNO ₃ :HF		143
ZnS	K ₂ Cr ₂ O ₇ :H ₂ SO ₄		143
AlN ^e	photoresist developer	0.01–1	145
GaN ^d	NaOH–H ₂ O	2	146
InN ^d	HCl–HNO ₃	1	146

^a GaAs selective to AlGaAs.

^b GaAs at 300 K.
^c GaAs agitated.
^d Rate nm min.
^e Active component is KOH.

Etch Profiles. The final profile of a wet etch can be strongly influenced by the crystalline orientation of the semiconductor sample. Many wet etches have different etch rates for various exposed crystal planes. In contrast, several etches are available for specific materials which show little dependence on the crystal plane, resulting in a nearly perfect isotropic profile. The different profiles that can be achieved in GaAs etching, as well as InP-based materials, have been discussed (130–132). Similar behavior can be expected for other crystalline semiconductors. It can be important to control the etch profile if a subsequent metallization step has to pass over the etched step. For reliable metal step coverage it is desirable to have a sloped etched step or at worst a vertical profile. If the profile is re-entrant (concave) then it is possible to have a break in the metal film, causing an open defect.

A powerful feature of wet etching is the ability to achieve excellent etch selectivities of one material over another. This can be extremely useful in the fabrication of epitaxial devices with different material layers. Because selective etching allows the removal of specific layers, the final accuracy of the etch can approach that of the epitaxial layers. Etch selectivities of >100 : 1 have been achieved for citric acid:H₂O₂ etching of GaAs–AlGaAs and InGaAs–InP structures (133).

Etch Mechanisms. Most wet etches for the compound semiconductors employ oxidation of the semiconductor followed by dissolution of the oxide. For this reason, many wet etches contain the oxidant hydrogen peroxide, although nitric acid can also be used. One advantage of wet etching over dry is the absence of subsurface damage that is common with dry etching. Metal contacts placed on wet-etched surfaces exhibit more ideal characteristics than dry-etched surfaces.

Etch Chemistries for Compound Semiconductors. There is a great diversity in the number of wet etches used for compound semiconductors. By choosing the correct etch the result can be highly selective or completely nonselective, etching through multiple epitaxial layers with nearly constant rates. The profile can be isotropic or depend on the crystal orientation. Most photoresists do a fairly good job of withstanding the wet etches. However, if a dielectric or metal surface is to be exposed during the etch it is important to consider parasitic etching of these materials (134). The large number of etchants available for compound semiconductors is extensive and therefore only a limited number of them are listed in Table 4. Several references for the wet etching of GaAs and InP are available (135,136).

DRY ETCHING

For certain applications dry etching has gained popularity over wet etching because of its increased control of etch profiles, attaining submicrometer features and the ability to introduce *in situ* monitoring capabilities into a dry-etch system. In general, dry etching of III–V semiconductors involves exposing the semiconductor to a directed energetic reactive plasma, which etches the semiconductor through a combination of physical and chemical processes. Different etch systems and or conditions can be used to vary the amount of physical or chemical etching. Ion-beam milling involves the purely physical sputtering action of argon atoms, while reactive ion etching (RIE) can have a very strong chemical component with highly isotropic profiles.

Dry-Etching Systems. The fundamental operational characteristic of a dry-etch system is the exposure of a semiconductor sample to energetic reactive gases. The gases that are introduced into the etch system have a large impact as to whether the etch has any chemical component or is a purely physical etch, as in ion milling. The pressures used are generally <13.3 Pa (0.1 mm Hg) in order to reduce gas-phase interactions, increase the mean free path of the energetic gases, and control lateral etching. Very anisotropic etching usually occurs with pressures <1.33 Pa and d-c biases >100 V. A plasma is generated in the etchant gases by several techniques, including radio-frequency (r-f) energy, microwave energy, and microwave energy combined with magnetic confinement. The plasma may be located remote from the sample or the sample may be directly within the plasma.

Ion-beam milling is a dry-etch process that is 100% physical, using argon plasmas in a highly energetic mode of operation. Chemically assisted ion-beam etching (CAIBE), reactive ion-beam etching (RIBE), and electron cyclotron resonance (ECR) etching are similar in that the plasmas are located remotely from the semiconductor. Systems that remove the sample from the plasma in a downstream mode allow excellent independent control of the physical and chemical components of the etch. Inductively coupled plasma (ICP) etching and reactive ion etching have the sample sitting directly in the plasma. However, ICP etching allows decoupling of the ion density and ion energy, resulting in excellent control over the etch characteristics.

The physical component of the etch can cause damage to the semiconductor surface (147). Although this damage can often be annealed out at low temperatures, this is not always possible in the fabrication sequence. Additionally, in some cases it is necessary to use a wet etch to remove a thin damaged layer formed during a dry etch. If the etch is designed to have a chemical component then the atomic species in the plasma interacts with the semiconductor forming volatile by-products which desorb from the surface. The sample stage can often be heated or cooled to control the volatility of these by-products. Etches with strong chemical components are more isotropic, but introduce less damage to the semiconductor surface.

As with any other fabrication process, masks are needed to define the features to be etched. It is common that the etch used for the semiconductor also etches the masking material. For this reason many different masks are used in etching, including photoresist, dielectric films, and metals. Masking can be a complex issue, especially when very deep etches (>5 μm) are performed with high aspect ratios (148).

Gases for Etching. The primary factor in choosing a gas for dry etching of compound semiconductors is its ability to form volatile compounds with the semiconductor's constituent atoms. These volatile compounds are either thermally desorbed from the surface or desorb through an ion-assisted process (149). All gases used have a physical component associated with its ability to sputter the semiconductor surface. A wide variety of gases have been used in the etching of compound semiconductors and listing them all is beyond the scope of this article. A short listing of gases for various semiconductors is shown in Table 5 (150,151).

ION IMPLANTATION

Although a great number of compound semiconductor devices make use of epitaxy to form the core vertical structure of the device, ion implantation (qv) is a powerful tool in creating both horizontal and vertical modifications to a device. Ion implantation can be used to dope a semiconductor either *n*- or *p*-type by using appropriate species. Implantation can also be used to render a region semi-insulating or to initiate multilayer intermixing.

Table 5. Dry-Etch Gases for Compound Semiconductors

Semiconductor	Etch method ^a (gas)	Reference
GaAs, AlGaAs	RIE (SiCl ₄ or Cl ₂)	152
	RIE (CH ₄ –H ₂)	153
	ECR (CH ₄ –H ₂ –Ar)	154
	ECR (HBr)	155
	ECR (CCl ₂ F ₂ , SiCl ₄ , or BCl ₃)	156
	ion milling (Ar)	147
	RIE (CH ₄ –H ₂)	157
InP	RIE (Cl ₂ –H ₂)	158
	ECR (CH ₄ –H ₂ –Ar)	159

InGaAs, InAlAs	ECR (Cl ₂ -H ₂)	160
	ECR (HBr)	155
	ion milling (Ar)	147
GaSb	ECR (HBr or CH ₄ -H ₂)	155,161
	RIE (Cl ₂ or SiCl ₄)	152
GaN, InN, AlN	hybrid ECR (CCl ₂ F ₂ -O ₂ or PCl ₂)	162
	ECR (Cl ₂ H ₂ , BCl ₃ , or CCl ₂ F ₂)	163,164
	ECR (HI-H ₂ -Ar or HBr-H ₂ -Ar)	165
HgCdTe, ZnS, CdTe	ECR (CH ₄ -H ₂ or H ₂ -Ar)	166,167

^a RIE = reactive ion etching; ECR = electron cyclotron resonance etching.

Ion Implantation Systems. An ion implantation system is used to accelerate ionized atomic or molecular species toward a target sample. The ionized species penetrates the surface of the sample with the resulting depth profile dependent on the implanted species mass, energy, and the sample target's tilt and rotation. An implanter's main components include an ionizer, mass separator, acceleration region, scanning system, and sample holder (168).

The implanted ion can be singly or multiply charged and can be any isotope. The mass separation system is used to avoid contamination. As an example, when implanting silicon the ²⁹Si⁺ isotope is often used instead of ²⁸Si⁺ to avoid contamination from the ²⁸N⁺₂ and ¹⁴N²⁺ signals. After mass separation the ions are electrostatically accelerated to their final energy, which can range from 5 keV to a few MeV. The deflector system and sample stage vary greatly system from system (169). Most implanters have some ability to tilt and rotate the sample with respect to the incoming beam. This is often used to avoid channeling of ions within the crystalline lattice of the semiconductor (170), achieving a more abrupt profile.

Implantation Basics. The depth distribution of the implanted species is a function of the acceleration energy, the ion mass, and the density and crystal structure of the sample (169). Two key parameters of the distribution are the projected range, ie, where the peak occurs, and the longitudinal straggle, which represents the spread of the implant. The implanted ion penetrates deeper into the sample for a higher implant energy, a lighter ion, or a less dense substrate. The depth distribution of the ions can be described by numerous analytical approximations including Gaussian and the more accurate Pearson IV models (169). The simpler Gaussian approximation has the following form, where $C(x)$

$$C(x) = \frac{D_0}{\Delta R_p \sqrt{2\pi}} \exp\left(-\frac{1}{2} \left(\frac{x - R_p}{\Delta R_p}\right)^2\right)$$

is the ion concentration in cm⁻³, R_p (cm) is the projected range, ΔR_p (cm) is the straggle, D_0 is the implanted dose in cm⁻², and x (cm) is the depth into the sample from the surface.

Many different materials can be used to spatially mask an implant on the semiconductor surface. Such masks include photoresist, dielectrics, and metals. In order to be an effective implant mask the material should be thick enough to prevent the implant from penetrating the mask and entering the sample. A minimum thickness for stopping 99.99% of the ions in the masking material is $R_p + 3.72\Delta R_p$ (168).

When the implanted species enters the crystalline semiconductor it is slowed by a combination of electronic and nuclear scattering events (169). The nuclear scattering events leave significant lattice damage, including vacancies, interstitials, and antisite defects that must be repaired before the sample is useful. This is accomplished by thermal annealing. The high temperature anneals allow the lattice atoms to reintegrate into the crystalline lattice. At the same time this annealing allows the implanted atoms to occupy lattice sites or become substitutional (171). This is often necessary for dopant atoms to become electrically active. If the implant is meant as an isolation step the annealing may be skipped as this might destroy the isolation. Treatment of the sample surface is critical during the anneal as the compound semiconductors often decompose at the high temperatures needed for activation (>600°C). Samples are often annealed with a dielectric cap to prevent decomposition or are placed face down on a sacrificial sample to induce a local Group V over pressure (171). Rapid thermal annealing (RTA) has been replacing furnace annealing as the predominant annealing technique. RTA has rapid rise and fall times during the temperature cycling which makes it easier to more consistently anneal samples with good activation and good surface control.

***n*- and *p*-Type Doping.** Ion implantation can be an effective technique for the selective incorporation of dopant atoms in a device structure. Although not possessing the atomic level precision of epitaxy, ion implantation is capable of defining submicrometer features in both vertical and horizontal directions. A wide variety of dopants have been used to dope the compound semiconductors.

The most common *n*-type dopants for the III-V semiconductors include silicon, sulfur, selenium, and tellurium (171). Silicon is the most popular as it suffers very little from diffusion during the anneal step following implantation, and generally has activation percentages over 70%. In general these dopants have good activation percentages except for high dose conditions. The heavier atoms (eg, tellurium) cause a great deal of lattice damage during implant, making it difficult to completely repair the lattice and causing a reduced activation percentage. Although silicon does not suffer from diffusion, it does suffer from amphoteric behavior. Being a Group IV element the silicon can occupy either the Group III or Group V lattice site. Proper annealing conditions, such as excess Group V and/or capping materials, are useful in ensuring high donor concentrations of silicon (171). The maximum *n*-type doping level achievable is approximately $2\text{--}4 \times 10^{18} \text{ cm}^{-3}$, which is lower than that for *p*-type doping (172).

For *p*-type doping in the III-Vs the most common dopants are beryllium, magnesium, zinc, carbon, and cadmium (171). The *p*-type dopants generally suffer from significant diffusion during the activation anneal (171). It is important to balance the anneal to achieve good activation without catastrophic diffusion. Beryllium generally displays excellent activation due to its light mass, which causes minimal lattice damage during implantation. Because the maximum acceptor concentration in the III-Vs is higher than that for donors, higher implantation doses can be used for *p*-type dopants. It is not unusual to achieve *p*-type dopant concentrations of 10^{20} cm^{-3} , although these samples often show serious diffusion. The diffusion of *p*-type dopants is often concentration dependent, getting worse with increasing concentration (173). Carbon is a *p*-type dopant that shows very little diffusion during annealing, but suffers from very poor activation. Coinplantation is often required to get acceptable activation levels with carbon implantation (174).

Ion implantation has been successfully used to dope the III-Sb material system. Sulfur has been used as an *n*-type dopant, although with poor activation efficiencies (175). *p*-Type doping has been achieved using beryllium, zinc, and magnesium (175,176). Activation of the *p*-type dopants is generally much better, near 50%. For the Sb-containing materials the post-implant anneal is conducted at much lower temperatures, typically <600°C.

Because of the relative immaturity of the III-N material systems and the difficulties associated with growing high quality films, only a small amount of ion implantation work has been carried out in this material system, including work on the physical properties of range and diffusion (177), isolation (178,179), and *n*- and *p*-type doping of GaN (180). Unlike the abundant work on ion implantation into the III-Vs, this technique has not been extensively used in the II-VI compound semiconductors. However, there are reports of work being done in doping HgCdTe (181) and other II-VI materials (182).

Isolation. Besides making semiconductor samples conducting, the other main use of ion implantation is to cause a region to be semi-insulating. This can be accomplished by two techniques: lattice damage or chemical modification. The damage caused by an implanted ion can cause a sample to become semi-insulating for the proper implant dose and energy. This is a temporary condition which can be reversed by thermal annealing (183). Typical ions that are used in this technique include hydrogen and oxygen. For very narrow band gap materials it is difficult to form good isolation regions due to thermal carrier generation.

To provide a more permanent isolation region an ion must be implanted that causes the sample to be semi-insulating after annealing. The most common ion used is oxygen (183), although iron has also been used in InP. Oxygen forms chemical compounds, which are deep trap states in compound semiconductors, depleting the samples of mobile charge. Deep trap states do not form in all semiconductors.

Intermixing. In some fabrication sequences it is desirable to cause the interdiffusion of semiconductor superlattices. This effect causes a

distinct two-material superlattice to interdiffuse, transforming into a single nearly homogeneous material. The optical and electrical properties of the new material is different from that of the superlattice. Extensive work has been done in the GaAs–AlGaAs material system using both silicon implantation (184) and proton implantation (185). Some work has also been carried out in other material systems such as InP–InGaAs (186,187) and InAlAs–InP (188).

METALLIZATION

Metals can be used in compound semiconductor processing for the formation of ohmic contacts, Schottky contacts, and interlevel wiring. There are vastly different requirements for each of these applications which has resulted in a large variety of metals being used in fabrication technologies. Good ohmic contacts should have low contact resistivity and good thermal stability. Good Schottky contacts should have clean interfaces with good barrier heights and thermal stability. Interlevel wiring should have low resistances with good adhesion.

Metal Deposition Systems. Metals can be deposited by a variety of techniques including thermal evaporation, electron beam evaporation, sputtering, and chemical vapor deposition (CVD). The most common of these techniques include thermal and electron beam evaporation. Both systems are identical except for the method used to heat the metal target. In each system the semiconductor sample and the metal target are located within a vacuum chamber that can be evacuated out to below 1.3×10^{-4} Pa (1×10^{-6} torr). The semiconductor sample surface is facing the metal target at a distance of usually 30 cm or more. There is a shutter between the metal target and the semiconductor to rapidly begin and end the deposition. The metal target is heated by one of two techniques: resistive or electron beam heating. The metal is evaporated from the target and deposited on the semiconductor surface. The amount of material deposited is monitored by a crystal oscillator. A common form of masking for evaporation is called lift-off (150). The masking material, often photoresist, is defined with a re-entrant profile that causes a break in the metal deposited on the semiconductor surface and the metal deposited on the masking surface. After metal deposition the masking material is chemically dissolved, lifting off the excess metal on the mask, leaving clearly defined metal regions on the semiconductor surface. Common metals used in thermal or electron beam evaporation include gold, titanium, platinum, palladium, nickel, germanium, and chrome.

Another common technique for metal deposition is sputtering (189). Sputtering is used for metals with high melting points that cannot easily be deposited by evaporation techniques or in situations where good step coverage is required. Additionally, reaction sputtering can also be used to form metal compounds such as TiN, by sputtering titanium in a nitrogen ambient (189). In a sputtering system a plasma is used to physically sputter the metal from a target which is then deposited on the semiconductor surface. Common metals for sputtering include tungsten, tungsten–silicide, titanium, tantalum, and transparent conductors such as indium tin oxide (ITO). Sputtering has a more isotropic deposition, making it difficult to use the lift-off technique. It is more common to employ post-deposition dry etching to define the metal.

The last technique commonly employed to deposit metals for compound semiconductors is electroplating (150). This technique is usually used where very thick metal layers are desired for very low resistance interconnects or for thick wire bond pads. Another common use of this technique is in the formation of air-bridged interconnects (150), which are popular for high speed electronic and optoelectronic circuits.

Metals for Ohmic Contacts. The main goal of an ohmic contact is to provide a low resistance contact to a heavily doped semiconductor region. Depending on the semiconductor to be contacted, the doping level, and the desired application, there are a number of different metallization schemes which can be used. In general it is easiest to make good ohmic contacts to heavily doped narrow band gap semiconductors. Additionally, as the alloy composition and band gap of a ternary or quaternary semiconductor changes, so does the quality of an ohmic contact made to the alloy. The most common form of contact used is the alloyed contact. Following metal deposition the contact and semiconductor are annealed to allow intermixing of the metal–semiconductor interface (190). This mixing introduces components of the metal into the semiconductor, often doping it to provide a low contact resistance contact. Table 6 lists a variety of contact metals, including a number of alloyed systems. An excellent list of alloyed contacts for GaAs is available (191). Many different metals have been used for a variety of semiconductors, and Table 6 lists only a small number of the possible combinations. The alloyed contact is not very shallow, because of mixing, and often has poor thermal stability because of continued interaction of the metals (190). Another approach is the nonalloyed contact, which displays ohmic behavior after deposition, requiring no further thermal processing. Unfortunately, this scheme cannot be used at all times. In general nonalloyed contacts require very high doping levels ($>10^{19}$ cm⁻³) and or low band gap semiconductors.

Table 6. Ohmic Contacts for Compound Semiconductors

Semiconductor doping, <i>n</i> or <i>p</i>	Metal system	Contact type ^a	Anneal temp, °C	Typical ρ_c , ^b Ωcm ²	Reference
<i>n</i> -GaAs	Au–Ni–Ge	A	400	10 ⁻⁶	192
	Ni–Ge–Au–W	A	650	1.3×10^{-6}	193
	Pd–Ge	SPR	250	1.5×10^{-6}	194
<i>p</i> -GaAs	Be–Au	A	525	1.2×10^{-7}	195
	Pd–Zn–Pd–Au	A	500	1×10^{-7}	196
<i>n</i> -InP	Au–Sn	A	400	2×10^{-6}	197
<i>p</i> -InP	Ge–Pd(Zn)	SPR	400	10 ⁻⁵	198
<i>p</i> -InAs	Pt–Ti	NA		7.5×10^{-6}	199
<i>n</i> -GaN	Ti–Al	A	900	8×10^{-6}	200
	Au	A	575	10 ⁻⁷	201
<i>n</i> -InN	Ti–Pt–Au	NA		2×10^{-7}	202
<i>n</i> -GaSb	Au–Te	A	400	10 ⁻⁶	203
<i>p</i> -GaSb	Au–Zn–Au	A	300	10 ⁻⁶	204
<i>p</i> -HgCdTe	Au	A	200	5×10^{-6}	205
<i>p</i> -ZnTe	Au–Pt–Ti–Ni	A	300	10 ⁻⁶	206

^a The contact type is alloyed, A; nonalloyed, NA; or solid-phase regrowth, SPR.

^b Typical contact resistivity

A third technique for ohmic contact formation is solid-phase regrowth (SPR) (190). SPR involves a modification of the deposited metals during thermal processing. The overall concept is that a thin epitaxial layer of the metal (usually Ge or Si) forms on the GaAs surface, providing a low resistance contact (190). The SPR contact is very shallow and can have good thermal stability. Although the SPR contact involves some post-deposition thermal processing, it does not suffer from the same metal–semiconductor mixing of the alloyed contact. Table 6 lists a number of SPR contacting recipes that have been successfully developed for compound semiconductors. One last technique involves the use of refractory metals in the ohmic contact. These contacts are useful for high temperature applications where the refractory metals do not melt or mix with the semiconductor. Although the contact resistances for these contacts are not as good as alloyed contacts, the improved thermal stability can be more important for specific applications.

Metals for Schottky Contacts. Good Schottky contacts on semiconductor surfaces should not have any interaction with the semiconductor as is common in ohmic contacts. Schottky contacts have clean, abrupt metal–semiconductor interfaces that present rectifying contacts to electron or hole conduction. Schottky contacts are usually not intentionally annealed, although in some circumstances the contacts need to be able to withstand high temperature processing and maintain good Schottky behavior.

The most common Schottky contacts for compound semiconductors are gold-based metallizations deposited by thermal or electron beam evaporation. The metal may include a thin titanium layer in direct contact with the semiconductor which acts as an adhesion layer. Additionally, a thin layer

of a diffusion barrier metal, such as platinum, can be placed between the titanium and the gold. Table 7 lists some standard Schottky contacts used for compound semiconductors. One problem with gold-based Schottky contacts are their poor thermal stability and inability to withstand high processing temperatures. The lift-off technique is often used in defining gold-based Schottky contacts.

Table 7. Schottky Barrier Heights for Metals on Compound Semiconductors

Semiconductor	Metal	Barrier height, eV	Reference
<i>n</i> -GaAs	Au	0.90	191
	W	0.65	191
<i>p</i> -GaAs	Au	0.42	44
<i>n</i> -InP	Au	0.52	44
<i>p</i> -InP	Au	0.76	44
<i>p</i> -InAs	Au	0.47	44
<i>p</i> -CdMnTe	Al	0.73	207
<i>n</i> -GaN	Au	0.91	208
	Ti	0.58	209
<i>n</i> -InSb	Au	0.05	210
<i>n</i> -GaSb	Al	0.65	211
<i>p</i> -AlSb	Au	0.71	212

A popular alternative to gold-based Schottky contacts is the use of refractory-based metals such as tungsten, tungsten silicide, or titanium nitride. These metals are deposited by sputtering and usually require post-deposition etching to define the contact region. These contacts usually have much better thermal stability and are often used in processes when high temperature anneals ($> 800^{\circ}\text{C}$) are used after contact formation. One drawback of using refractory metals is the high resistance of the metal as compared to gold-based contacts. This resistance can have deleterious effects on device performance.

Metals for Interlevel Wiring. The most common metals for interlevel wiring on compound semiconductors are gold, titanium–gold, and aluminum. The metals are used to connect individual devices into circuits. The metals are often required to travel over interlayer dielectric films, making contact to both Schottky and ohmic contact regions. The interlevel wiring should have low resistance for thin wires and should not suffer from any reactions with dielectrics, metals, or the semiconductor. Although aluminum is cheap and widely used in silicon processing it is not often used in high speed compound semiconductor circuits due to its resistance and known stability problems (168). Gold-based wiring is more common because of its excellent conductance and nonreactivity.

DIELECTRIC OVERLAYERS

Unlike silicon processing where metal oxide semiconductor (MOS) devices are a crucial technology, dielectric films in compound semiconductor processing have generally played a secondary role. There has been a great deal of research in trying to develop metal insulator semiconductor (MIS) devices for compound semiconductors, but successes have been fairly limited. A more common use of dielectric insulators is in encapsulation, planarization, interlevel dielectric spacers, and surface protection during annealing.

Dielectric Deposition Systems. The most common techniques used for dielectric deposition include chemical vapor deposition (CVD), sputtering, and spin-on films. In a CVD system thermal or plasma energy is used to decompose source molecules on the semiconductor surface (189). In plasma-enhanced CVD (PECVD), typical source gases include silane, SiH₄, and nitrous oxide, N₂O, for deposition of silicon nitride. The most common CVD films used are silicon dioxide, silicon nitride, and silicon oxynitrides.

Insulator sputtering is similar to the process described for metal sputtering. The only difference is that the source target is a dielectric film. There is less control of the chemical nature and quality of the film as compared to a CVD deposited film. Common sputtered films include silicon nitride and silicon oxide.

A number of dielectric films are deposited by the spin-on technique. In this case the film's constituent molecules are dissolved in a solvent to form a liquid. After spinning the liquid over a semiconductor surface the solvent is driven off with a baking step, leaving behind the thin dielectric film. Common films include polyimide and benzocyclobutene (BCB). The deposition process for these films is simple, making it attractive for a manufacturing process.

Dielectrics for Metal Insulator Semiconductor Structures. The biggest problem in using dielectric films for MIS applications is the poor quality of the interface between insulator and semiconductor. Typically this interface is characterized by a high density of traps which prevents good modulation of the carrier density in the semiconductor. The primary cause of these traps is a large density of dangling bonds on the semiconductor surface. In order to fabricate a high quality MIS structure, passivation of these dangling bonds is necessary. Some of the most common techniques for this include the use of sulfur passivation, either by aqueous or plasma techniques (213). Sufficient progress has been made in the development of MIS structures that basic MISFETs are being fabricated in InP, GaAs, and InGaAs (213,214). Some attempts have also been made in using AlN as an insulator layer for GaAs-based MIS devices (215).

Dielectrics for Interlevel Wiring and Planarization. When connecting a number of devices to form a circuit it is often necessary to cross wires without actual electrical contact between them. In order to accomplish this multiple levels of wiring separated by dielectric films are necessary. In some cases it is possible to form bridges of metal using air as the dielectric (150). However, if more than two levels of wiring are required then dielectric spacing is necessary. The ideal dielectric film has excellent adhesion and a low dielectric constant to minimize parasitic capacitances. The most common films include silicon oxide, silicon nitride, and a number of spin-on dielectrics (216).

During the fabrication process the surface of the semiconductor is etched and metal contacts are deposited. These features can represent a topographical challenge to subsequent metal wiring levels. For this reason it is important that the dielectric film used tends to smooth out such discontinuities as metal and etched edges (150,217). Additional applications for spin-on dielectrics include forming integrated microlenses for optoelectronics (218).

LITHOGRAPHY

There are a number of different techniques for pattern definition on a semiconductor sample, however, none are particular to compound semiconductors. The main goal of lithography is to accurately define a region to be modified, either by deposition, etching, or masking of an implant. The region to be modified is often defined by its location to previous features and therefore the accurate alignment of a mask to previous features is a significant component of lithography.

The most common method for the definition of a pattern is optical lithography. Optically sensitive resists are exposed to light which affects their dissolution rate when exposed to a developing solution. Resists that etch away when exposed to uv light are referred to as positive tone, whereas resists that become etch resistant when exposed to light are referred to as negative tone. In general, positive tone resists provide higher resolution and are more useful in a number of fabrication steps. Techniques for masking the resist include contact printing and projection printing (150). In contact printing, a glass plate with a patterned metal mask is put in contact with the resist-coated sample and light is shown through the mask. In a projection system, the mask is held a fixed distance from the sample and light is focused through the mask and onto the sample. Projection systems can include pattern reduction optics after the mask. Typically these systems employ stepping of the semiconductor wafer, exposing a single die at a time. Both techniques are in use, with stepper projection systems more popular in production facilities. The resolution of optical techniques is basically limited by the wavelength of the light used

to expose the mask. Modern optical systems accurately define features of 0.25-μm resolution, and are getting smaller annually.

To achieve smaller dimensions there are systems that use x-rays instead of optical photons (150). These systems require a collimated x-ray source which is often expensive. Additionally, systems have been developed that use ion beams to expose the resist. Either an x-ray or ion beam system requires specialized resists and exposure systems.

A maskless lithographic technique, commonly referred to as *e*-beam lithography, uses high energy steerable electron beams (150). These systems operate with very narrow beams of electrons having energies ranging from 25 to 100 keV, depending on the particular system. The electron beam is electrostatically scanned around the sample surface, exposing the resist to the electrons. The electrons can cause the resist to either cross-link or break down depending on the exact chemistry of the resist. The noncross-linked resist is then etched away in a development process. Typical electron beam lithography systems can achieve minimum feature dimensions <50 nm. They are popular in the production of masks for both optical and x-ray techniques as well as for direct writing on semiconductor surfaces. However, the low throughput of *e*-beam systems makes them presently intractable for production use in most circumstances. They are particularly popular in research laboratories exploring nanoscale (<0.1 μm) devices.

This work is supported by the U.S. Department of Energy under contract #DE-AC04-94AL85000.

BIBLIOGRAPHY

"Semiconductors Fabrication" in *ECT* 2nd ed., Vol. 17, pp. 862–883, by J. R. Carruthers, Bell Telephone Labs, Inc.; in *ECT* 3rd ed., Vol. 20, pp. 634–654, by C. C. Chang, Bell Laboratories; "Semiconductors, Theory and Application" in *ECT* 3rd ed., Vol. 20, pp. 601–633, S. A. Schwartz, Bell Laboratories, Inc.

1. S. Strite and H. Morkoz, *J. Vac. Sci. Technol. B* **10**, 1237 (1992).
2. H. Morkoz, S. Strite, G. B. Gao, M. E. Lin, B. Sverdlov, and M. Burns, *J. Appl. Phys.* **76**, 1363 (1994).
3. H. Luo and J. K. Furdyna, *Semicond. Sci. Technol.* **10**, 1041 (1995).
4. J. Brice and P. Capper, eds., *Properties of Mercury Cadmium Telluride*, Institution of Electrical Engineers (INSPEC), London, 1987.
5. R. K. Willardson and A. C. Beer, eds., *Mercury Cadmium Telluride*, Academic Press, Inc., New York, 1981.
6. J. K. Furdyna, *J. Appl. Phys.* **64**, R29 (1988).
7. G. B. Stringfellow, *Organometallic Vapor-Phase Epitaxy: Theory and Practice*, Academic Press, Inc., New York, 1989.
8. R. F. C. Farrow, ed., *Molecular Beam Epitaxy: Applications to Key Materials*, Noyes Publications, Park Ridge, N.J., 1995.
9. S. M. Hu, *J. Appl. Phys.* **69**, 7901 (1991).
10. G. C. Osbourn, *J. Vac. Sci. Technol. B* **1**, 379 (1983).
11. A. Baldereschi and N. O. Lipari, *Phys. Rev. B* **3**, 439 (1971).
12. A. Baldereschi and N. O. Lipari, *Phys. Rev. B* **8**, 2697 (1973).
13. O. Madelung, M. Schulz, and H. Weiss, eds., *Landolt-Burnstein*, Vol. 17, *Semiconductors*, Springer-Verlag, New York, 1985.
14. Ref. 13, 1987.
15. INSPEC, *Electronic Materials Information Series*, Institution of Electrical Engineers, London.
16. T. F. Kuech, *Mater. Sci. Rep.* **2**, 1 (1987).
17. P. D. Dapkus, *Ann. Rev. Mat. Sci.* **12**, 243 (1982).
18. *J. Crystal Growth* **145** (1994), for examples.
19. D. W. Shaw, in C. H. L. Goodman, ed., *Crystal Growth: Theory and Techniques*, Academic Press, Inc., New York, 1974.
20. T. J. Mountziaris and K. F. Jensen, *J. Electrochem. Soc.* **138**, 2426 (1991).
21. J. Van De Ven, G. M. J. Rutten, M. J. Raaijmakers, and L. J. Giling, *J. Crystal Growth* **76**, 352 (1986).
22. K. F. Jensen, D. I. Fotiadis, and T. J. Mountziaris, *J. Crystal Growth* **107**, 1 (1991).
23. W. G. Breiland and G. H. Evans, *J. Electrochem. Soc.* **138**, 1806 (1991).
24. D. W. Kisker, G. B. Stephenson, P. H. Fuoss, and F. J. Lamelas, *J. Crystal Growth* **124**, 1 (1992).
25. J. E. Butler, N. Bottka, R. S. Stillman, and D. K. Gaskill, *J. Crystal Growth* **77**, 163 (1986).
26. G. A. Hebner, K. P. Killeen, and R. M. Biefeld, *J. Crystal Growth* **98**, 293 (1989).
27. D. E. Aspnes, *J. Crystal Growth* **120**, 71 (1992).
28. K. P. Killeen and W. G. Breiland, *J. Electron. Mater.* **23**, 179 (1994).
29. B. R. Butler and J. P. Stagg, *J. Crystal Growth* **94**, 293 (1989).
30. A. C. Jones, *J. Crystal Growth* **129**, 728 (1993).
31. T. F. Kuech, E. Veuhoff, T. S. Kuan, V. Deline, and R. Potemski, *J. Crystal Growth* **77**, 257 (1986).
32. R. P. Schneider, R. P. Bryan, E. D. Jones, R. M. Biefeld, and G. R. Olbright, *J. Crystal Growth* **123**, 487 (1992).
33. G. Haake and S. P. Watkins, *J. Crystal Growth* **107**, 342 (1991).
34. B. Y. Maa and P. D. Dapkus, *J. Electron. Mater.* **20**, 589 (1991).
35. B. A. Banse and J. R. Creighton, *Appl. Phys. Lett.* **60**, 856 (1992).
36. F. J. Lamelas, P. H. Fuoss, P. Imperatori, D. W. Kisker, and G. B. Stephenson, *Appl. Phys. Lett.* **60**, 2610 (1992).
37. H. Asai, *J. Crystal Growth* **80**, 425 (1987).
38. D. M. Kozuch, M. Stavola, S. J. Pearton, C. R. Abernathy, and W. S. Hobson, *J. Appl. Phys.* **73**, 3716 (1993).
39. E. T. J. M. Smeets, *J. Crystal Growth* **82**, 385 (1987).
40. D. M. Follstaedt, R. M. Biefeld, S. R. Kurtz, and K. C. Baucom, *J. Electron. Mater.* **24**, 819 (1995).
41. J. M. Vandenberg, R. A. Hamm, and S. N. G. Chu, *J. Crystal Growth* **144**, 9 (1994).
42. M. W. Wang, D. A. Collins, T. C. McGill, R. W. Grant, and R. M. Feenstra, *J. Vac. Sci. Technol. B* **13**, 1689 (1995).
43. S. M. Sze, *Semiconductor Devices: Physics and Technology*, John Wiley & Sons, Inc., New York, 1985.
44. S. M. Sze, *Physics of Semiconductor Devices*, 2nd ed., John Wiley & Sons, Inc., New York, 1981.
45. M. Shur, *GaAs Devices and Circuits*, Plenum Press, New York, 1987.
46. W. T. Read, *Bell Syst. Tech. J.* **37**, 401 (1958).
47. W. Mickanin, P. Canfield, E. Finchem, and B. Odekirk, in *Proceedings of the GaAs IC Symposium*, San Diego, Calif., 1989.
48. T. Ando, *Rev. Mod. Phys.* **54**, 437 (1982).
49. S. M. Sze, *High-Speed Semiconductor Devices*, John Wiley & Sons, Inc., New York, 1990.
50. M. Wojtowicz and co-workers, *IEEE Elec. Dev. Lett.* **15**, 477 (1994).
51. S. Luryi, *IEEE Trans. Elec. Dev.* **41**, 2241 (1994).
52. K. Kurishima, T. Kobayashi, and U. Gosele, *Appl. Phys. Lett.* **60**, 2496 (1992).
53. J. J. Liou, *Advanced Semiconductor Device Physics and Modeling*, Artech House, Boston, Mass., 1994.
54. T. Nittono, H. Ito, O. Nakajima, and T. Ishibashi, *Jpn. J. Appl. Phys.* **27**, 1718 (1988).
55. S. J. Pearton, *Int. J. Mod. Phys.* **7**, 4687 (1993).
56. H. Kroemer, *Proc. IEEE* **70**, 13 (1982).
57. R. T. Bate, *Sci. Am.* **258**, 78 (1988).
58. J. N. Randall, M. A. Reed, and G. A. Frazier, *J. Vac. Sci. Technol. B* **7**, 1398 (1989).
59. A. C. Seabaugh, J. H. Luscombe, and J. N. Randall, *Fut. Elec. Dev. J. (Jpn.)* **3**, 9 (1993).
60. J. N. Randall, *Nanotechnology* **4**, 41 (1993).

61. M. A. Reed and W. P. Kirk, eds., *Nanostructure Physics and Fabrication*, Academic Press, Inc., New York, 1989.

62. C. S. Lent, P. D. Tougaw, W. Porod, and G. H. Bernstein, *Nanotechnology* **4**, 49 (1993).

63. J. H. Luscombe, *Nanotechnology*, **4**, 1 (1993).

64. J. R. Soderstrom, D. H. Chow, and T. C. McGill, *IEEE Elec. Dev. Lett.* **11**, 27 (1990).

65. E. R. Brown, J. R. Soderstrom, C. D. Parker, L. J. Mahoney, K. M. Molvar, and T. C. McGill, *Appl. Phys. Lett.* **58**, 2291 (1991).

66. C. E. Chang, P. M. Asbeck, K. C. Wang, and E. R. Brown, *IEEE Trans. Elec. Dev.* **40**, 685 (1993).

67. R. C. Potter, A. A. Lakhani, and H. Hier, *J. Appl. Phys.* **64**, 3735 (1988).

68. A. C. Seabaugh, Y. C. Kao, and H. T. Yuan, *IEEE Elec. Dev. Lett.* **13**, 479 (1992).

69. A. A. Lakhani and R. C. Potter, *Appl. Phys. Lett.* **52**, 1684 (1988).

70. L. P. Kouwenhoven, A. T. Johnson, N. C. van der Vaart, D. J. Maas, C. J. P. M. Hammans, and C. T. Foxon, *Surf. Sci.* **263**, 405 (1992).

71. G. Zimmedi, T. M. Eiles, R. L. Kautz, and J. M. Martinis, *Appl. Phys. Lett.* **61**, 237 (1992).

72. H. Grabert and M. H. Devoret, eds., *Single Charge Tunneling: Coulomb Blockade Phenomena in Nanostructures*, Plenum Press, New York, 1992.

73. A. N. Korotkov, R. H. Chen, and K. K. Likharev, *J. Appl. Phys.* **78**, 2520 (1995).

74. K. Yano, T. Ishii, T. Hashimoto, T. Kobayashi, F. Murai, and K. Seki, *IEEE Trans. Elec. Dev.* **41**, 1628 (1994).

75. B. E. A. Saleh and M. C. Teich, *Fundamentals of Photonics*, John Wiley & Sons, Inc., New York, 1991.

76. C. Yeh, *Applied Photonics*, Academic Press, Inc., New York, 1994.

77. Y. Suematsu and A. R. Adams, eds., *Handbook of Semiconductor Lasers and Photonic Integrated Circuits*, Chapman and Hall, New York, 1994.

78. D. Wood, *Optoelectronic Semiconductor Devices*, Prentice Hall, New York, 1994.

79. S. Nakamura, *J. Cryst. Growth* **145**, 911 (1994).

80. D. B. Eason and co-workers, *Appl. Phys. Lett.* **66**, 115 (1995).

81. J. W. Cook, Jr. and J. F. Schetzina, *Laser Focus World* **31**, 101 (1995).

82. M. G. Crawford, *IEEE Circuits Dev.* **8**, 24 (1992).

83. K. Itaya, H. Sugawara, and G. Hatakoshi, *J. Cryst. Growth* **138**, 768 (1994).

84. F. A. Kish and co-workers, *Appl. Phys. Lett.* **64**, 2839 (1994).

85. E. F. Schubert, Y.-H. Wang, A. Y. Cho, L.-W. Tu, and G. J. Zydzik, *Appl. Phys. Lett.* **60**, 921 (1992).

86. M. S. Unlu and S. Strite, *J. Appl. Phys.* **78**, 607 (1995).

87. J. A. Lott, J. Schneider, J. C. Zolper, and K. J. Malloy, *IEEE Photon. Technol. Lett.* **5**, 631 (1993).

88. N. E. J. Hunt, E. F. Schubert, R. F. Kopf, D. L. Sivco, A. Y. Cho, and G. J. Zydzik, *Appl. Phys. Lett.* **63**, 2600 (1993).

89. T. R. Chen, L. Eng, Y. H. Zhang, A. Yariv, N. S. Kwong, and P. C. Chen, *Appl. Phys. Lett.* **56**, 1345 (1990).

90. O. Mikami, H. Yasaka, and Y. Noguchi, *Appl. Phys. Lett.* **56**, 987 (1990).

91. S. Kondo, H. Yasaka, Y. Noguchi, K. Magari, M. Sugo, and O. Mikami, *Electron. Lett.* **28**, 132 (1992).

92. J. Hecht, *The Laser Guidebook*, 2nd ed., McGraw Hill Book Co., Inc., New York, 1992.

93. T. V. Higgins, *Laser Focus World* **31**, 65 (1995).

94. P. S. Zory, ed., *Quantum Well Lasers*, Academic Press, Inc., San Diego, Calif., 1993.

95. A. Yariv, *Quantum Electronics*, 3rd ed., John Wiley & Sons, Inc., New York, 1989.

96. G. A. Evans and J. M. Hammer, eds., *Surface Emitting Semiconductor Lasers and Arrays*, Academic Press, Inc., San Diego, Calif., 1993.

97. S. Nakamura and co-workers, *Jpn. J. Appl. Phys.* **35**, L217 (1996).

98. R. L. Gunshor and A. V. Nurmikko, *Laser Focus World* **31**, 97 (1995).

99. M. A. Haase, J. Qui, J. M. DePuydt, and H. Cheng, *Appl. Phys. Lett.* **59**, 1272 (1991).

100. S. Itoh and A. Ishibashi, in R. L. Gunshor and A. V. Nurmikko, eds., *Proc. SPIE* **2346**, 2 (1994).

101. J. M. Gaines, R. R. Drenten, K. W. Haberern, T. Marshal, P. Mensz, and J. Petruzzello, *Appl. Phys. Lett.* **63**, 2315 (1993).

102. M. A. Haase and co-workers, *Appl. Phys. Lett.* **63**, 2315 (1993).

103. D. C. Grillo and co-workers, *Appl. Phys. Lett.* **63**, 2723 (1993).

104. D. C. Grillo and co-workers, *Electron. Lett.* **30**, 2131 (1994).

105. H. Jeon and co-workers, *Electron. Lett.* **31**, 106 (1995).

106. P. D. Floyd, J. L. Merz, H. Luo, J. K. Furdyna, T. Yokogawa, and Y. Yamada, *Appl. Phys. Lett.* **66**, 2929 (1995).

107. D. P. Bour, D. W. Treat, K. J. Beernik, B. S. Krusor, R. S. Geels, and D. F. Welch, *IEEE Photon. Technol. Lett.* **6**, 128 (1994).

108. J. A. Skidmore and co-workers, *IEEE Photon. Technol. Lett.* **7**, 133 (1995).

109. J. A. Lott, R. P. Schneider, K. D. Choquette, S. P. Kilcoyne, and J. J. Figiel, *Electron. Lett.* **29**, 1693 (1993).

110. M. Hagerott Crawford, J. Schneider, K. D. Choquette, and K. L. Lear, *IEEE Photon. Technol. Lett.* **7**, 724 (1995).

111. M. Razeghi, *Nature* **369**, 631 (1994).

112. K. L. Lear, K. D. Choquette, J. Schneider, S. P. Kilcoyne, and K. M. Geib, *Electron. Lett.* **31**, 208 (1995).

113. G. M. Yang, M. H. MacDougall, and P. D. Dapkus, *Electron. Lett.* **31**, 886 (1995).

114. Y. Hayashi, T. Mukaihara, M. Hatori, N. Ohnoki, A. Matsutani, F. Koyama, and K. Iga, *Electron. Lett.* **31**, 560 (1995).

115. K. L. Lear, J. Schneider, K. D. Choquette, S. P. Kilcoyne, J. J. Figiel, and J. C. Zolper, *IEEE Photon. Technol. Lett.* **6**, 1053 (1994).

116. R. A. Morgan and co-workers, *Electron. Lett.* **31**, 462 (1995).

117. B. Bouchert, R. Gessner, and B. Stegmuller, *Jpn. J. Appl. Phys.* **33**, 1034 (1994).

118. T. Baba, Y. Yogo, K. Suzuki, F. Koyama, and K. Iga, *Electron. Lett.* **29**, 913 (1993).

119. D. I. Babic, J. J. Dudley, K. Streubel, R. P. Mirin, J. E. Bowers, and E. L. Hu, *Appl. Phys. Lett.* **66**, 1030 (1995).

120. Z. Feit, K. Kostyk, R. J. Woods, and P. Mak, *Appl. Phys. Lett.* **58**, 343 (1991).

121. H. K. Choi, G. W. Turner, and S. J. Eglash, *IEEE Photon. Technol. Lett.* **6**, 7 (1994).

122. H. K. Choi and G. W. Turner, *Appl. Phys. Lett.* **67**, 332 (1995).

123. P. N. J. Dennis, *Photodetectors: An Introduction to Current Technology*, Plenum Press, New York, 1986.

124. L. J. Kozlowski, J. M. Arias, G. M. Williams, K. Vural, D. E. Cooper, S. A. Cabelli, and C. Bruce, in R. E. Longshore, ed., *Proc. SPIE* **2274**, 93 (1994).

125. P. R. Norton, in Ref. 124, p. 82.

126. S. Gunapala, G. Sarusi, J. Park, T.-L. Lin, and B. Levine, *Phys. World* **7**, 53 (1994).

127. B. F. Levine, A. Zussman, S. D. Gunapal, M. T. Asom, J. M. Kuo, and W. S. Hobson, *J. Appl. Phys.* **72**, 429 (1992).

128. A. Rogalski, in E. L. Dereniak and R. E. Sampson, eds., *Proc. SPIE* **2225**, 118 (1994).

129. A. Srinivasan, S. Murtaza, J. C. Campbell, and B. G. Streetman, *Appl. Phys. Lett.* **66**, 535 (1995).

130. D. W. Shaw, *J. Electrochem. Soc.* **128**, 874 (1981).

131. T. Takebe, T. Yamamoto, M. Fujii, and K. Kobayashi, *J. Electrochem. Soc.* **140**, 1169 (1993).

132. A. Stano, *J. Electrochem. Soc.* **134**, 448 (1987).

133. G. C. DeSalvo, W. F. Tseng, and J. Comas, *J. Electrochem. Soc.* **139**, 831 (1992).

134. J. L. Vossen and W. Kern, eds., *Thin Film Process*, Academic Press, London, 1978.

135. *Properties of Gallium Arsenide*, INSPEC, London, 1990.

136. *Properties of InP*, INSPEC, London, 1991.

137. S. Iida and K. Ito, *J. Electrochem. Soc.* **118**, 768 (1971).

138. D. W. Shaw, *J. Electrochem. Soc.* **128**, 874 (1981).

139. R. A. Logan and F. K. Reinhart, *J. Appl. Phys.* **44**, 4172 (1973).

140. R. Becker, *Solid-State Electron.* **16**, 1241 (1973).

141. S. Uekusa, K. Oigawa, and M. Yacano, *J. Electrochem. Soc.* **132**, 671 (1985).

142. G. C. DeSalvo, R. Kaspi, and C. A. Bozada, *J. Electrochem. Soc.* **141**, 3526 (1994).

143. B. Tuck, *J. Mater. Sci.* **10**, 321 (1975).

144. M. Illing, G. Bacher, A. Forchel, T. Litz, A. Waag, and G. Landwehr, *Appl. Phys. Lett.* **66**, 1815 (1995).

145. J. R. Mileham, S. J. Pearton, C. R. Abernathy, J. D. MacKenzie, R. J. Shul, and S. P. Kilcoyne, *Appl. Phys. Lett.* **67**, 1119 (1995).

146. S. J. Pearton, C. R. Abernathy, F. Ren, J. R. Lothian, P. W. Wisk, and A. Katz, *J. Vac. Sci. Technol. A* **11**, 1772 (1993).

147. S. J. Pearton, U. K. Chakrabarti, and A. P. Perley, in A. Katz, R. M. Biefeld, R. L. Gunshor, and R. J. Malik, eds., *Proc. Mater. Res. Soc.* **216**, 507 (1990).

148. J. R. Lothian, F. Ren, and S. J. Pearton, *Semiconductor Sci. Technol.* **7**, 1199 (1992).

149. D. M. Manos and D. L. Flamm, eds., *Plasma Etching: An Introduction*, Academic Press, Inc., New York, 1989.

150. R. E. Williams, *Gallium Arsenide Processing Techniques*, Artech House, Inc., Dedham, Mass., 1984.

151. A. Katz, ed., *Indium Phosphide and Related Materials: Processing, Technology and Devices*, Artech House, Boston, Mass., 1992.

152. S. J. Pearton, U. K. Chakrabarti, W. S. Hobson, and A. P. Kinsella, *J. Vac. Sci. Technol. B* **8**, 607 (1990).

153. R. Cheung, S. Thomas, S. P. Beamont, G. Doughty, V. Law, and C. D. W. Wilkinson, *Elec. Lett.* **231**, 857 (1987).

154. S. J. Pearton, U. K. Chackrabarti, A. P. Perley, and W. S. Hobson, *J. Electrochem. Soc.* **138**, 1432 (1991).

155. S. J. Pearton, U. K. Chakrabarti, E. Lane, A. P. Perley, C. R. Abernathy, and W. S. Hobson, *J. Electrochem. Soc.* **139**, 856 (1992).

156. S. J. Pearton, *Mat. Sci. Eng. B* **10**, 187 (1991).

157. T. R. Hayes, M. A. Dreisbach, P. M. Thomas, W. C. Dautremont-Smith, and L. A. Heimbrook, *J. Vac. Sci. Technol. B* **7**, 1130 (1989).

158. G. A. Vawter and C. I. H. Ashby, *J. Vac. Sci. Technol. B* **12**, 3374 (1994).

159. S. J. Pearton, U. K. Chakrabarti, A. P. Kinsella, D. Johnson, and C. Constantine, *Appl. Phys. Lett.* **56**, 1424 (1990).

160. C. Constantine, C. Barratt, S. J. Pearton, F. Ren, and J. R. Lothian, *Electron. Lett.* **28**, 1749 (1992).

161. S. J. Pearton, U. K. Chakrabarti, W. S. Hobson, C. Constantine, and D. Johnson, *Nuc. Instr. Meth. Phys. Res. B* **59–60**, 1015 (1991).

162. S. J. Pearton, U. K. Chakrabarti, A. Katz, A. P. Perley, W. S. Hobson, and M. Geva, *Plas. Chem. Plas. Proc.* **11**, 405 (1991).

163. R. J. Shul, S. P. Kilcoyne, M. Hagerott Crawford, J. E. Parmeter, C. B. Vartuli, C. R. Abernathy, and S. J. Pearton, *Appl. Phys. Lett.* **66**, 1761 (1995).

164. S. J. Pearton, C. B. Vartuli, R. J. Shul, and J. C. Zolper, *Mat. Sci. Eng. B* **31**, 309 (1995).

165. S. J. Pearton, C. R. Abernathy, and C. B. Vartuli, *Elec. Lett.* **30**, 1985 (1994).

166. C. R. Eddy, Jr., E. A. Dobisz, J. R. Meyer, and C. A. Hoffman, *J. Vac. Sci. Technol. A* **11**, 1763 (1993).

167. S. J. Pearton and F. Ren, *J. Vac. Sci. Technol. B* **11**, 15 (1993).

168. S. K. Ghandhi, *VLSI Fabrication Principles*, John Wiley & Sons, Inc., New York, 1983.

169. H. Ryssel and H. Glawischnig, eds., *Ion Implantation Techniques*, Springer-Verlag, Berlin, 1982.

170. R. T. Blunt and P. Davies, *J. Appl. Phys.* **60**, 1015 (1986).

171. S. J. Pearton, *Int. J. Mod. Phys. B* **7**, 4687 (1993).

172. S. J. Pearton, W. S. Hobson, and C. R. Abernathy, in N. W. Cheung, A. D. Marwick, and J. B. Roberto, eds., *Proc. Mat. Res. Soc.* **147**, 261 (1989).

173. M. Uematsu, K. Wada, and U. Güsele, *Appl. Phys. A* **55**, 301 (1992).

174. B. P. Davies, P. Davies, D. M. Brookbanks, D. J. Warner, and R. H. Wallis, *International Symposium on GaAs and Related Compounds*, Jersey, Channel Islands, U.K., 1990.

175. M. V. Rao, P. E. Thompson, R. Echard, S. Mulpuri, A. K. Berry, and H. B. Dietrich, *J. Appl. Phys.* **69**, 4228 (1991).

176. A. G. Milnes and co-workers, *Mater. Sci. Eng. B* **27**, 129 (1994).

177. R. G. Wilson, C. B. Vartuli, C. R. Abernathy, S. J. Pearton, and J. M. Zavada, *Solid-States Electron.* **38**, 1329 (1995).

178. J. C. Zolper, S. J. Pearton, C. R. Abernathy, and C. B. Vartuli, *Appl. Phys. Lett.* **66**, 3042 (1995).

179. S. C. Binari, H. B. Dietrich, G. Kelner, L. B. Rowland, K. Doverspike, and D. K. Wickenden, *J. Appl. Phys.* **78**, 3008 (1995).

180. S. J. Pearton, C. B. Vartuli, J. C. Zolper, C. Yuan, and R. A. Stall, *Appl. Phys. Lett.* **67**, 1435 (1995).

181. B. L. Sharma and R. Nath, *Solid State Data A* **80**, 1 (1991).

182. F. J. Bryant, *Prog. Cryst. Growth Char.* **6**, 191 (1983).

183. S. J. Pearton, *Mater. Sci. Rep.* **4**, 313 (1990).

184. P. Chen and A. J. Steckl, *J. Appl. Phys.* **77**, 5616 (1995).

185. G. F. Redinbo, H. G. Craighead, and J. Minghuang Hong, *J. Appl. Phys.* **74**, 3099 (1993).

186. P. G. Piva and co-workers, *Superlattices/Microstruct.* **15**, 385 (1994).

187. W. Xia and co-workers, *J. Appl. Phys.* **71**, 2602 (1992).

188. S. Yamamura, T. Kimura, R. Saito, S. Yugo, M. Murata, and Y. Kamiya, in H. S. Rupprecht and G. Weimann, eds., *Proceedings of the 20th International Symposium on Gallium Arsenide and Related Compounds*, Freiburg, Germany, 1993, p. 485.

189. S. M. Rossmagel, J. J. Cuomo, and W. D. Westwood, eds., *Handbook of Plasma Processing Technology—Fundamentals, Etching, Deposition, and Surface Interactions*, Noyes Publications, Park Ridge, N.J., 1990.

190. L. J. Brillson, eds., *Contacts to Semiconductors—Fundamentals and Technology*, Noyes Publications, Park Ridge, N.J., 1993.

191. M. J. Howes and D. V. Morgan, eds., *Gallium Arsenide—Materials, Devices and Circuits*, John Wiley & Sons, Inc., New York, 1986.

192. C. Lin and C. P. Lee, *J. Appl. Phys.* **67**, 260 (1990).

193. N. Lustig, M. Murakami, M. Norcott, and K. McGann, *Appl. Phys. Lett.* **58**, 2093 (1991).

194. M. L. Lovejoy, A. J. Howard, K. R. Zavadil, D. J. Rieger, R. J. Shul, and P. A. Barnes, *J. Vac. Sci. Technol. A* **13**, 758 (1995).

195. R. Fischer and H. Morkoc, *IEEE Elec. Dev. Lett.* **EDL-7**, 359 (1986).

196. R. Bruce, D. Clark, and S. Eicher, *J. Elec. Mater.* **19**, 225 (1990).

197. P. A. Barnes and R. S. Williams, *Solid-State Electron.* **24**, 907 (1981).

198. L. C. Wang, M. H. Park, F. Deng, A. Clawson, S. S. Lau, D. M. Hwang, and C. J. Palmstrom, *Appl. Phys. Lett.* **66**, 3310 (1995).

199. A. Katz and co-workers, *J. Vac. Sci. Technol. B* **8**, 1125 (1990).

200. M. E. Lin, Z. Ma, F. Y. Huang, Z. F. Fan, L. H. Allen, and H. Morkoc, *Appl. Phys. Lett.* **64**, 1003 (1994).

201. J. S. Foresi and T. D. Moustakas, *Appl. Phys. Lett.* **62**, 2859 (1993).

202. F. Ren, C. R. Abernathy, S. N. G. Chu, J. R. Lothian, and S. J. Pearton, *Appl. Phys. Lett.* **66**, 1503 (1995).

203. E. Villemain, S. Gaillard, M. Rolland, and A. Joullie, *Mat. Sci. Eng. B* **20**, 162 (1993).

204. J. B. B. Oliveira, C. A. Olivieri, J. C. Galzerani, A. A. Pasa, and F. C. de Prince, *J. Appl. Phys.* **66**, 5484 (1989).

205. K. Kim, H. Chung, S. C. Kim, H. C. Lee, C. K. Kim, H. K. Kim, and J. M. Kim, *J. Korean. Inst. Telematics Electron.* **31A**, 87 (1994).

206. K. Mochizuki, A. Terano, M. Momose, A. Taike, M. Kawata, J. Gotoh, and S. Nakatsuka, *Elec. Lett.* **30**, 1984 (1994).

207. J. Szatkowski, E. Placzek-Popko, B. Bieg, A. Hajdusianek, and S. Kuzminski, *17th International Seminar on Surface Physics*, Kudowa, Poland, 1995.

208. M. R. H. Khan, T. Detchprohm, P. Hacke, K. Hiramatsu, and N. Sawaki, *J. Phys. D* **28**, 1169 (1995).

209. S. C. Binari, H. B. Dietrich, G. Kelner, L. B. Rowland, K. Doverspike, and D. K. Gaskill, *Elec. Lett.* **30**, 909 (1994).

210. M. McColl and M. F. Millea, *J. Elec. Mater.* **5**, 191 (1976).

211. A. Y. Polyakov, M. Stam, A. G. Milnes, and T. E. Schlesinger, *Mat. Sci. Eng. B* **12**, 337 (1992).

212. S. Sadiq and A. Joullie, *J. Appl. Phys.* **65**, 4924 (1989).
213. D. S. L. Mui, Z. Wang, and H. Morkoc, *Thin Solid Films* **231**, 107 (1993).
214. J. Reed and co-workers, *Solid-State Electron.* **38**, 1351 (1995).
215. F. Alexander, J. M. Masson, G. Post, and A. Scavennec, *Thin Solid Films* **98**, 75 (1982).
216. P. B. Chonoy and J. Tajadod, *IEEE Trans. Comp. Hybrids Man. Tech.* **16**, 714 (1993).
217. L. K. White, *J. Electrochem. Soc.* **130**, 1543 (1983).
218. O. Blum and co-workers, *Elec. Lett.* **3**, 44 (1995).

Timothy J. Drummond
Robert M. Biefeld
Marc E. Sherwin
Mary H. Crawford
Sandia National Laboratories



SENSORS

Sensor technology is being revolutionized by advances in microelectronics and optoelectronics. The newer sensors are demonstrating improvements in the operation, reliability, safety, and efficiency of many engineering systems. Perhaps the most familiar consumer example of this technology is found in automobiles where silicon-micromachined pressure sensors, coupled to an electrochemical oxygen sensor, allow the engine management computer to provide the correct fuel and spark parameters for smooth, low pollution operation. Some automobiles include rotation rate sensors for antilock brakes, fluid flow sensors for gas mileage computation, accelerometers for air bag deployment, and temperature sensors for closed-loop temperature control of the passenger area. Development of other sensors for safety, comfort, and pollution control are underway.

Increasingly, the word sensor is used interchangeably with, or in place of, the word transducer to represent the conversion of one type of energy into another. The word actuator, however, refers to a distinctly different set of devices capable of activating a device upon receiving an appropriate signal. Sensors and actuators may be found as part of a system on a single piece of silicon.

Sensor systems of the 1990s almost always include data acquisition capabilities. Because of the low cost of computers, microprocessors, analog digital converting circuit boards, and other peripherals, the concept of large arrays of distributed sensors has become commonplace. The sensor itself is often the weak link in such a system, partly because the investment in sensor technology has been dwarfed by the development of microelectronics technology, but also because of fundamental differences in the nature of the problem. Personal computers deal with a relatively limited set of physical variables, ie, principally the flow of electrons through semiconductors and metals. With the exception of some impressively precise and sophisticated electromechanical engineering for disk drives and printers, the means of computer input and output are relatively minor extensions of extremely well-developed technologies: the typewriter keyboard, the cathode ray tube, and sometimes audio speakers.

In contrast, various sensors are expected to respond in a predictable and controlled manner to such diverse parameters as temperature, pressure, velocity or acceleration of an object, intensity or wavelength of light or sound, rate of flow, density, viscosity, elasticity, and, perhaps most problematic, the concentration of any of millions of different chemical species. Furthermore, a sensor that responds selectively to only a single one of these parameters is often the goal, but the first attempt typically produces a device that responds to several of the other parameters as well. Interferences are the bane of sensors, which are often expected to function under, and be immune to, extremely difficult environmental conditions.

Although it is possible to obtain, manipulate, display, and act on huge amounts of data, the degree to which such data are meaningful often hinges on the reliability of the sensor itself. The pervasiveness of inexpensive microprocessors has also freed sensor technology to exploit phenomena that give logarithmic and other nonlinear responses to measurands. In some commercial instruments, the calibration curves for a large number of individual sensors are stored, so that the instrument knows the characteristics of each sensor sold and can give the correct values for the measurand. One example of this capability is pH measurement. One problem not solved as of 1996 by computer technology (qv) is a sensor output that is faithfully repeatable. Automated or automatic calibration is a partial solution to this problem, however.

In the field of chemical sensors, the revolution in software and inexpensive hardware means that not only nonlinear chemical responses can be tolerated, but incomplete selectivity to a variety of chemical species can also be handled. Arrays of imperfectly selective sensors can be used in conjunction with pattern recognition algorithms to sort out classes of chemical compounds and their concentrations when the latter are mixed together.

Decision-Making Tool for Sensor Technology

The rapid advances in sensor technology have increased the difficulty of the task of finding the best sensor system for a given application. Figure 1 shows a flow chart designed to be used as a decision-making tool for sensor technology (1). The sensor customer or engineer needs to be specific about defining the sensor problem to be solved, using language and descriptors that are commonly shared between the customer and a sensor technologist. An extensive sensor glossary and set of sensor descriptors is available (1). A sensor technologist or a sensor buyer's guide such as the annual guide produced by *Sensors Magazine* (2) can be a great help in locating vendors for various technologies.

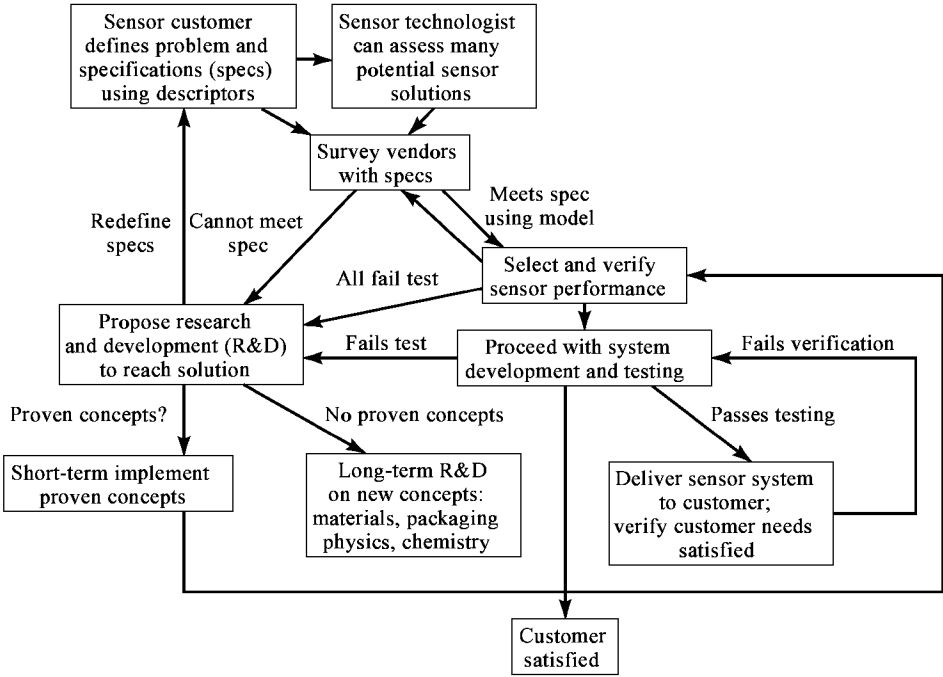


Fig. 1. Decision-making tool for sensor technology (1).

Many physical principles can be employed by different sensor vendors to obtain the same measurement. For example, in 1995, 49 vendors were listed for hydrogen sensing (2). Represented among the sensor systems are mass spectrometers, gas chromatographs, electrochemical cells, thermal

conductivity cells, Raman scattering instruments, metal-oxide semiconductors, and pellistor (catalytic material-coated wire) sensors. To the process of sensor selection, the engineer brings specifications such as cost, size, weight, sensitivity, selectivity, and lifetime. The engineer must also specify the environmental conditions to which the sensor system is to be subjected: high or low temperature, vacuum, corrosive chemicals, vibration, etc. The specifications can be cast in the form of descriptors and matched against each vendor's performance specifications. No common database for performance specifications is available as of this writing (ca 1996), and often the literature for each vendor must be compiled at a cost of much time and effort. Performance under hostile conditions is often not available, forcing the sensor engineer into a test program to verify the commercial sensor's performance. In many cases, all the commercially available sensor systems fail to meet the specifications, leaving the engineer with the choice of redefining the specifications or launching a development project to fabricate a sensor system that solves the problem.

Developing a new solution may take the form of a modification of sensors already available or using the sensor principles of an available sensor, but modifying it to meet the specifications. On the other hand, new principles and manufacturing techniques may be explored in a long-range project to yield improved sensor performance. There is a large and growing research and development (R&D) effort in sensor technology because many engineers see that the advances in microelectronics, optoelectronics, and biotechnology are pointing the way to sensor performance unobtainable with commercially available sensor systems.

An important part of the decision-making tool flow chart (see Fig. 1) is the redefinition of the specifications by the sensor customer after the surveying and testing of commercial sensor systems. This process almost always involves a downgrading of the expectations of the sensor customer, but it can also force a realistic evaluation of what sensor information really needs to be made available and at what cost.

In an era of inexpensive, batch-processed sensors, the concept of throwing away the sensor element, ie, the transducer, actually becomes cost-effective. An example would be the miniature electrochemical cells on a paper test strip for monitoring blood glucose. The instrument for measuring the electrochemical current may be relatively expensive, but each test strip costs only pennies, in spite of the apparent complexity of having screen-printed electrodes with a membrane containing dry glucose oxidase. Under long-time exposure to blood, this sensor no longer gives the calibrated electrochemical current level for values stored in the reader memory. Thus the sensor, not adequate for closed-loop control of an implanted insulin pump, is adequate for a quick, one-time reading of glucose in a fresh drop of blood.

Another important path on the flow diagram (see Fig. 1) is the short-term implementation of proven concepts. Often it is the packaging, not the sensor itself, that fails in the test environment. An old sensor may be improved by application of more rugged packaging, encapsulation, wiring, etc. When all else fails, however, the customer may have to fund longer-term directed research and development on unproven sensor concepts. The strategy for this kind of R&D involves a strongly interdisciplinary activity, combining materials science, electrical engineering, chemistry, packaging science, etc (1).

A good survey of many sensors can be found in the literature (3,4). A number of international journals are also devoted to new sensor technology (see *General References*). Sensor papers often appear in journals such as *Analytical Chemistry* (including a biannual review of chemical sensors), *Journal of the Electrochemical Society*, and *The Journal of Applied Physics*. Several international conferences focus on sensors, most notably the biannual *Transducers: International Conference on Solid-State Sensors and Actuators*. Digests from the *Transducers* conferences are very good references to the state of the art. The Society of Photooptical Instrumentation Engineers (SPIE) conducts many sensor symposia and publishes proceedings volumes.

Sensors Based on Silicon Processing Technology

The art of batch processing has brought the silicon-based integrated circuit (IC) (see INTEGRATED CIRCUITS) and related electronic components to an almost unbelievably low cost. A digital watch, which could not be purchased at any price in 1970, could be had for as little as \$1.49 retail in 1995. The watch contains integrated circuits, a quartz crystal oscillator, a liquid crystal (LC) display, a battery, and the case. For a somewhat higher price, watches having sensors for temperature, barometric pressure, altitude, and compass direction were available.

The tools for the revolution in new sensors include micromachining, batch processing, integrated sensors, readout and networking electronics, light-emitting and detecting structures, fiber optics (qv), etc. This revolution has led and is leading to new applications of old fields of science from mechanics to physical chemistry, fluid mechanics, electrokinetics, capillary action, etc. These principles are being applied in a micro nano-scale world, where batch processing means inexpensive, rugged systems (see NANOTECHNOLOGY (SUPPLEMENT)).

Photodetectors. The most commercially successful silicon-based sensors are the photodetectors (qv). These are found in video cameras, security systems, ionizing radiation detectors, and many other applications. Photodetectors, closely related to solar cells in the principles of operation, all depend on light and ionizing radiation creating electron-hole pairs in the semiconducting material. There are many ways that the presence of the excess number of electron-hole pairs can be monitored. The simplest is the observation of a change in electrical conductivity, ie, photoconductivity. Ingenious schemes involving the storage of photoelectrons or holes are used to make the very sensitive charge-coupled device (CCD) arrays.

There are also a large number of photodetectors made from semiconductors other than silicon. Gallium arsenide, GaAs, and its alloys are the most common because the wavelengths to which they are sensitive can be widely varied, and light-emitting sources such as light-emitting diodes (LEDs) and lasers (qv) can be made from the same materials (see LIGHT GENERATION, LIGHT EMITTING DIODES). The compact-disk player is a good example of the common use of both a semiconductor photoemitter and detector.

Micromachines. One of the fastest-growing areas of sensors based on silicon involves the use of micromachining to fabricate the sensor. The largest commercial markets are for micromachined pressure sensors and accelerometers. Chemical etching of the silicon is the key technology for the fabrication of these sensors. Isotropic etching refers to a process in which the etch rate of the materials is uniform in all directions. Anisotropic etching refers to etch rates that depend on the crystallographic orientation of the silicon wafer. Anisotropy ratios of 400:1 are possible for special directions. Doping of the silicon can have a large effect on the etch rate, and layers of different materials such as SiO₂ and Si₃N₄ can have different etch rates. For pressure sensors, thin diaphragms of Si or related materials are etched into the wafer (see PRESSURE MEASUREMENTS).

Many of the variations developed to make pressure sensors and accelerometers for a wide variety of applications have been reviewed (5). These sensors can be made in very large batches using photolithographic techniques that keep unit manufacturing costs low and ensure part-to-part uniformity. A pressure differential across these thin diaphragms causes mechanical deformation that can be monitored in several ways: piezoresistors implanted on the diaphragm are one way; changes in electrical capacitance are another.

Smart Sensors. A smart sensor is loosely defined as a sensing device with built-in intelligence that usually involves some kind of digital signal processing. Because the sensor is already being batch-fabricated on a silicon wafer, it seems natural to create the signal-processing electronics on the same chip as the sensor itself. Some devices based on this technology are commercially available. An advanced example is given in Figure 2 (5). This pressure sensor is based on an anisotropically etched cavity, four piezoresistors which occupy the large square at the center of the chip, surrounded by digital electronics that correct the offset and sensitivity of temperature coefficients of the signal from the sensor along with other parameters. The calibration points are stored in a 30-bit static shift register, and the electronics needed for communicating the pressure values to a small computer are also integrated on the chip.

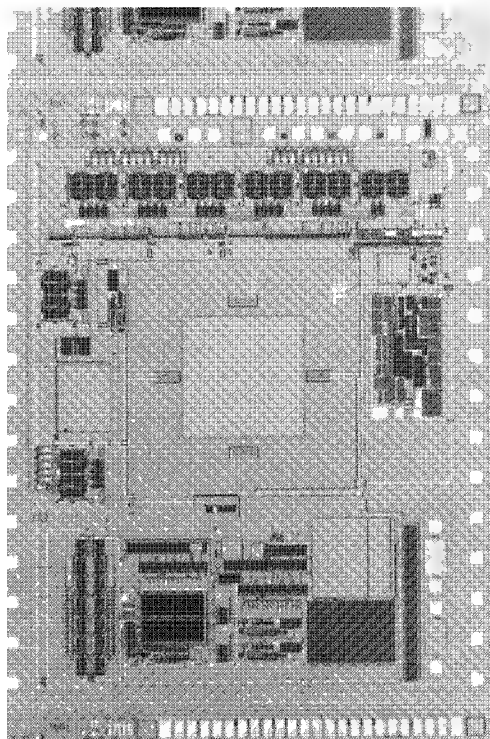


Fig. 2. Digital pressure sensor on a 6×4.8 mm chip, fabricated at the Technical University of Berlin (5). See text.

Pressure sensors that give temperature-corrected, linear, analog voltage output are available from Motorola and other manufacturers. In such sensors, the on-chip electronics correct any temperature effects and nonlinearities in the output of the piezoresistors. The on-chip electronics replace a shoebox-size collection of printed circuit boards. The price of this kind of smart sensor is considerably less than \$100. The integration of a large amount of circuitry on the chip allows functions like amplification, offset correction, self-testing, autocalibration, interference reduction, and compensation of cross-sensitivities (6).

The integration of standard integrated circuitry and a sensor is not always a trivial extension of the IC process, because the processing steps involved in fabricating the sensor may be incompatible with the correct functioning of the IC. An alternative is the hybrid package where the sensor chip and the signal-processing chip are fabricated separately and then glued together in the same package. There are a number of methods for connecting the leads from the sensor chip to the signal-processing IC. The hybrid method may be more expensive to package, but the integrated smart sensor may entail expensive and time-consuming process development to implement the ultimately less expensive final chip. Whereas sophisticated smart sensors like the one shown in Figure 2 may appear to be a manufacturable device, the sensor may not be available for purchase by the sensor customer until a vendor steps forward and takes the responsibility for manufacturing, packaging, servicing, and sales.

Even the tight controls in silicon integrated circuit manufacturing are not yet sufficient to produce absolutely identical sensors on a single wafer. Calibration of the final product is usually necessary, often by adjusting the value of a circuit element on the IC such as a resistor. The calibration process can be automated, but it still adds to the cost of batch-fabricated sensors. Clever means of self-calibration, particularly in field use, are constantly being sought.

Chemical Sensors. Silicon microelectronics are often touted as giving the most sensitive and stable measurements of physical phenomena, eg, charged particle detectors that can record a single x-ray photon or beta particle. In the realm of chemical detection, biological systems unequivocally outperform all of the microelectronically based chemical sensors in terms of sensitivity and selectivity. One of the best-studied examples of biological chemical detection is the sex attractant receptor of moths. Single-molecule detection of the sex attractant molecule, bombykol, can be claimed if the firing of a single neuron by probes is monitored. However, if the response of the moth is the criterion, an airstream containing 1000 molecules cm^3 or 1 part in 10^1 at standard temperature and pressure is required (7). The moth's microsensor is a receptor on a membrane protein that is part of the ion channel in the nerve cell wall.

A silicon-based sensor that operates on a similar principle is the chemically sensitive field-effect transistor (ChemFET) (8–10). The channel through which the electrical current flows is a very thin (about 10-nm) layer in the silicon crystal near a planar interface with an oxide of silicon that is thermally grown on the crystal. This interface has received an enormous amount of interest because it forms the basis for most active devices in integrated circuits. The very high sensitivity of the ChemFET is in part a result of it acting as a chemical amplifier. In these small devices, as few as 10 hydrogen atoms occupying interfacial sites can cause an order of magnitude increase in conductivity (10).

ChemFETs have been fabricated for the detection of a wide variety of chemical species. In the gas phase, catalytic metal gates (field plates) have been used to sense hydrogen, ammonia, ethanol, formic acid, ethylene, carbon monoxide, and many other molecules. Ion-conductive liquids, usually water, can also form the field plate, and pH-sensitive dielectric layers have led to commercially available pH sensors (see HYDROGEN ION ACTIVITY). Other ions can also be detected by placing various kinds of ion-sensitive membranes on the dielectric. Detectors known as biosensors (qv) have been fabricated by placing membranes containing immobilized enzymes on the dielectric. These devices are often called enzyme-immobilized FET (ENFETs). An often-studied ENFET is the one that detects glucose. It usually works by immobilizing the enzyme glucose oxidase. The presence of glucose causes the production of gluconic acid in the membrane, and the pH change is detected by the FET (8–10).

Technologies Other Than Silicon

Acoustic Wave Sensors. Another emerging physical transduction technique involves the use of acoustic waves to detect the accumulation of species in or on a chemically sensitive film. This technique originated with the use of quartz resonators excited into thickness-shear resonance to monitor vacuum deposition of metals (11). The device is operated in an oscillator configuration. Changes in resonant frequency are simply related to the areal mass density accumulated on the crystal face. These sensors, often referred to as quartz crystal microbalances (QCMs), have been coated with chemically sensitive films to produce gas and vapor detectors (12), and have been operated in solution as liquid-phase microbalances (13). A dual QCM that has one smooth surface and one textured surface can be used to measure both the density and viscosity of many liquids in real time (14).

Another class of acoustic wave sensors uses surface acoustic waves (SAWs) to probe the accumulation of species in a surface film. These devices use interdigital electrodes that are photolithographically formed on the surface of a piezoelectric crystal to excite and detect the surface wave. The stress-free boundary condition imposed by the surface results in a mode of propagation that is confined to the surface and has acoustic energy distributed within one wavelength of the surface. This surface confinement of acoustic energy makes the SAW extremely sensitive to the properties of a surface film, particularly changes in its mass. The surface wave has two components of displacement, one normal to the surface and one in the direction of wave propagation. It is the translation of surface mass by these SAW displacement components that leads to a change in surface wave velocity. Mass sensitivity can be further

enhanced by propagating a wave in a very thin membrane (15), generating an acoustic mode known as a Lamb wave.

The acoustic wave gravimetric technique can be extended to liquid-phase sensing in one of two ways. Because the SAW’s surface-normal displacement component generates compressional waves in a contacting liquid, excessive attenuation of the SAW occurs, rendering it unsuitable for liquid-phase measurements. A mode having exclusively in-plane surface displacement, such as the shear-horizontal acoustic plate mode, can, however, be used successfully (16). Shear waves propagate poorly in liquids, so attenuation is minor. The quartz thickness-shear resonators described also have no surface-normal displacement and thus function in contact with liquids. A second alternative is to use an acoustic mode of velocity slower than that of sound in the contacting liquid, such as the Lamb wave, which is described in detail in the literature (15).

A number of gas and vapor sensors have been realized by placing chemically sensitive films on SAW devices and relying on their extreme mass sensitivity. The first such sensor, reported in 1979 (17), detected a range of organic vapors utilizing thin films of vacuum grease, oils, and waxes commonly used in gas chromatography (qv). More than 100 papers have been published in this field since that time, including several reviews (18,19). Some of the most intensely studied areas include the use of organic polymer films to detect organic solvent vapors (20–23); monitoring small inorganic molecules, such as the halogens, NO_x, and sulfur oxides, using organometallic films, for example, metal phthalocyanines (24,25) or metal oxides (26); humidity sensors based on organic polymer films (27,28); and the detection of organophosphonates, simulants for and decomposition products of chemical warfare agents (29,30), using a variety of different thin-film materials. Styrene has been detected by its reversible reaction with an organoplatinum compound (31).

Microsensors Based on Optical Fibers. Optical fibers offer a number of advantages for chemical and physical sensing (32). This technology offers long-distance telemetry, freedom from electromagnetic interference, and small size, ie, approximately that of a human hair. Furthermore, the basic components necessary for sensor applications are being developed by the optical communications industry. Indeed, optical communications systems and optical sensors have the same functional design. Both need a light source, a detector, and a method of transmitting the light from the source and to the detector, namely, the optical fiber. For optical communications, a modulator is used to impose the information to be transmitted on the light beam. For optical sensing, the sensing element replaces the modulator and again imposes information on the light beam.

The sensing element can modify either the intensity, phase, or polarization state of the light, and physical sensors have been made that make use of all three techniques (33,34). For chemical sensing, most transduction mechanisms have focused on intensity modulation. Spectroscopy and fluorescence-based fiber optic sensors are in this group. A small effort (35) has been aimed at utilizing phase-modulating sensors, ie, interferometers. The development of fiber optic chemical sensors has been slower than other chemical and physical sensing techniques. Research in the field performed by members of the Society of Photooptical Instrumentation Engineers (SPIE) is recorded in their annual Chemical, Biochemical, and Environmental Fiber Sensors symposium.

Electrochemical Microsensors. The most successful chemical microsensor in use as of the mid-1990s is the oxygen sensor found in the exhaust system of almost all modern automobiles (see EXHAUST CONTROL, AUTOMOTIVE). It is an electrochemical sensor that uses a solid electrolyte, often doped ZrO₂, as an oxygen ion conductor. The sensor exemplifies many of the properties considered desirable for all chemical microsensors. It works in a process-control situation and has very fast (~100 ms) response time for feedback control. It is relatively inexpensive because it is designed specifically for one task and is mass-produced. It is relatively immune to other chemical species found in exhaust that could act as interferants. It performs in a very hostile environment and is reliable over a long period of time (36).

The success of the O₂ sensor has made the auto manufacturers, regulators, and environmentalists anxious to extend chemical sensing to a variety of tailpipe gases, notably CO, NO_x, and short-chain hydrocarbons. Considerable research and development is needed for these molecules to be monitored in the hostile exhaust system environment (36).

Ion-selective electrodes and amperometric cells have had a long history of success in a wide variety of applications (8,9). A microelectronics-inspired revolution is also occurring in these devices, brought about by the advent of photolithographically defined arrays of microelectrodes on planar substrates (37).

Semiconducting Metal Oxide Sensors. Another kind of conductometric sensor that is commercially available and has growing sales is formed from ceramic semiconducting metal oxides. The most common type of sensor consists of a thin porous film of tin oxide, SnO₂, deposited with a heater and electrodes to monitor the metal oxide conductivity. These sensors are often called Taguchi gas sensors, after the inventor. They are primarily used in small stand-alone sensor packages for detecting hazardous gases such as carbon monoxide and natural gas, ie, methane and butane, leaks. The mechanism for the change of conductivity in the presence of these reducing gases involves oxidative and reductive chemical reactions at the exposed surface and near-surface regions of the metal oxide. The dopant that makes these wide band gap materials semiconducting is actually a missing oxygen in the lattice, ie, a vacancy. Because of the porous and heterogeneous nature of the ceramic films, there is still considerable debate about the roles of various surface reactions in the sensing response. Typically, the ceramic must be quite hot (>200° C) to operate quickly and reversibly. The history, principles of operation, and performance data for the tin oxide-based sensor have been described (38). Many other metal oxide semiconductors have been investigated for their gas sensing properties, and a vast literature exists describing many preparation procedures, the addition of catalysts, the integration of the metal-oxide film on silicon, etc. This literature shows a large development effort to overcome the shortcomings of this type of sensor, which include poor selectivity, in which some home CO detectors based on this technology can sound false alarms for nail polish, various polishes and cleaners, and ethanol; large power consumption, in which the same home CO sensors are invariably powered from the a-c power lines, not batteries; problems with consistency in manufacturing; lack of a stable baseline; etc (38).

Economic Aspects

Sensors form a very broad-based, multibillion dollar business. Detailed information and predictions of the growth in the many sensor subfields can be found in the literature (39). For example, the relatively narrow area of acceleration and vibrations sensors was a \$600 million business in 1995 and projected to become a \$1 billion business by the end of the twentieth century. New applications, often driven by regulatory and safety concerns, mean projected growth of just about every sensor type.

Sensors based on microelectronics and photonics advances, along with gains in materials and interfacial science, are certain to provide higher performance sensor systems. The feasibility of much higher performance in terms of size, power consumption, and selectivity than that available as of the 1990s is demonstrated by biological sensors. The revolution in microelectronics manufacturing technologies nevertheless shows a growing trend toward diminutive cost as well as size. In the particularly difficult case of chemical sensing, considerable development is expected to be necessary in order to match the sensing abilities of an animal, such as a dog, to pick just one familiar example, but progress in that direction is expected.

ACKNOWLEDGMENT

This article was written at Sandia National Laboratories with support provided by the U.S. Department of Energy under Contract No. DE AC04 94AL85000.

BIBLIOGRAPHY

1.

Expanding the Vision of Sensor Materials, National Materials Advisory Board (NMAB) Publication 470, National Academy Press, Washington, D.C, Apr. 1995.

2.

Sensors 1995 Buyers Guide, Helmers Publishing Inc., St. Peterborough, N.H.

3.

W. Gupel, J. Hesse, and J. N. Zemel, eds., *Sensors: A Comprehensive Survey*, VCH Publishers, New York, 1989.

4.

J. Fraden, *AIP Handbook of Modern Sensors*, American Institute of Physics, New York, 1993.

5. J. Bryzek, K. Petersen, and W. McCulley, *IEEE Spectrum*, 20–31 (May 1994).

6. S. Middelhoek, *Sensors/Actuators* **A41–42**, 1–8 (1994).

7. D. Schneider, *Sci. Am.* **231**, 28 (July 1974).

8. J. Janata, *Principles of Chemical Sensors*, Plenum Publishing Corp., New York, 1989.

9. M. J. Madou and S. R. Morrison, *Chemical Sensing with Solid State Devices*, Academic Press, Inc., New York, 1989.

10. R. C. Hughes, A. J. Ricco, M. A. Butler, and S. J. Martin, *Science* **254**, 74–80 (1991).

11. G. Z. Sauerbrey, *Z. Physik.* **155**, 206 (1959).

12. T. Nakamoto and T. Morizumi, *Proc. IEEE Ultrason. Symp.*, 613 (1988).

13. T. Numura and A. Minemura, *Nippon Kagaku Kaishi*, 1621 (1980).

14. S. J. Martin, G. C. Frye, and K. O. Wessendorf, *Sensors/Actuators* **A44**, 209–218 (1994).

15. S. W. Wenzel and R. M. White, *Sensors/Actuators* **A21–23**, 700 (1990).

16. S. J. Martin, A. J. Ricco, T. M. Niemczyk, and G. C. Frye, **20**, 253 (1989).

17. H. Wohltjen and R. Dessy, *Anal. Chem.* **51**, 1465 (1979).

18. D. S. Ballantine, Jr., and H. Wohltjen, *Anal. Chem.*, **61**, 704A (1989).

19. M. S. Nieuwenhuizen and A. Venema, *Sensors Mater.* **1**, 261 (1989).

20. H. Wohltjen, A. W. Snow, W. R. Barger, D. S. Ballantine, *IEEE Trans. Ultrason., Ferroelec., Freq. Ctrl.* UFFC-34, 172 (1987).

21. J. W. Grate and co-workers, *Anal. Chem.* **60**, 869 (1988).

22. W. M. Heckl, F. M. Marassi, K. M. R. Kallury, D. C. Stone, and M. Thompson, *Anal. Chem.* **62**, 32 (1990).

23. G. C. Frye and S. J. Martin, *Sensors Mater.* **2**, 187 (1991).

24. M. S. Nieuwenhuizen and A. J. Nederlof, *Anal. Chem.* **60**, 230 (1988).

25. A. W. Snow, W. R. Barger, M. Klusty, and H. Wohltjen, *Langmuir* **2**, 513 (1986).

26. J. F. Vetelino, R. K. Lade, R. S. Falconer, *IEEE Trans. Ultrason., Ferroelec., Freq. Ctrl.* UFFC-34, 156 (1987).

27. J. G. Brace, T. S. Sanfelippo, S. G. Joshi, *Sensors Actuators* **14**, 47 (1988).

28. S. J. Martin, G. C. Frye, A. J. Ricco, and T. E. Zipperian, *Proc. IEEE Ultrasonics Symp.*, 563 (1987).

29. J. Lint, H. Wohltjen, N. L. Jarvis, and A. W. Snow, in D. Schuetzle and R. Hammerle, eds., *Fundamentals and Applications of Chemical Sensors*, ACS Symposium Series, No. 309, American Chemical Society, Washington, D.C., 1986, p. 155.

30. J. W. Grate, A. Snow, D. S. Ballantine, and H. Wohltjen, *Anal. Chem.* **60**, 864 (1988).

31. E. T. Zellers, N. Hassold, R. M. White, and S. M. Rappaport, *Anal. Chem.*, **62**, 1227 (1990).

32. C. Yeh, *Handbook of Fiber Optics, Theory and Applications*, Academic Press, New York, 1990.

33. J. Dakin and B. Culshaw, *Optical Fiber Sensors*, Vol. I, *Principles and Components*, Artech House, London, 1988.

34. J. Dakin and B. Culshaw, *Optical Fiber Sensors*, Vol. II, *Systems and Applications*, Artech House, London, 1989.

35. R. A. Lieberman and M. T. Wlodarczyk, eds., "Chemical, Biochemical, and Environmental Fiber Sensors II," *Proceedings of SPIE*, 1991, p. 1368.

36. E. M. Logothetis, in N. Yamazoe, ed., *Chemical Sensor Technology*, Vol. 3, Elsevier Science, Inc., New York, 1991, 89.

37. M. Morita, M. L. Longmire, and R. W. Murray, *Anal. Chem.* **60**, 2770 (1988).

38. K. Ihokura and J. Watson, *The Stannic Oxide Gas Sensor*, CRC Press, Ann Arbor, Mich., 1994.

39. *Sensor Bus. Digest*, Foster City, Calif., a monthly publication.

General References

Sensor Business Digest, monthly publication of Vital Information Publications, Foster City, Calif.

Sensor Magazine, monthly publication of Helmers Publishing Inc., St. Peterborough, N.H.

Sensors and Actuators, monthly publication of Elsevier Science, Inc., New York.

Sensors and Materials, bimonthly publication of the Scientific Publishing Division of MYU, Tokyo, Japan.

R. C. Hughes
A. J. Ricco
M. A. Butler
S. J. Martin
Sandia National Laboratories

SEPARATION

Centrifugal separation,
Magnetic separation,
Size separation,

CENTRIFUGAL SEPARATION

Centrifugal separation is a mechanical means of separating the components of a mixture of liquids or of liquids and solid particles. The material is accelerated in a centrifugal field which acts upon the mixture in the same manner as a gravitational field. The centrifugal field can, however, be varied by changes in rotational speed and equipment dimensions, whereas gravity is essentially constant. Commercial centrifugal equipment can reach an acceleration of 20,000 times gravity (20,000 *G*); laboratory equipment can reach up to 360,000 *G*. Most centrifugation equipment is intended to separate immiscible or insoluble components from a liquid medium. The ultracentrifuge and gas centrifuge represent special cases that establish separation gradients on a molecular scale. The usual gravitational operations, such as sedimentation (qv) or flotation (qv) of solids in liquids, drainage or squeezing of liquids from solid particles, and stratification of liquids according to density, are accomplished more effectively in a centrifugal field (see SEPARATION, SIZE SEPARATION). The development of theory for centrifugation equipment has been slow. Flow patterns in the centrifuge bowls are complex and difficult to model mathematically. The concept of a theoretical capacity factor for sedimentation depends only on equipment characteristics and is independent of the system (1,2). The theoretical effect of particle size distribution in single and multistage centrifugation has been demonstrated (3). Extensive application of centrifugation to dewatering (qv) and thickening of relatively soft solids and hydrogels associated with industrial and municipal waste treatment has resulted in changes in centrifuge design (see WASTES, INDUSTRIAL). A sound theoretical basis for centrifugal drainage or squeezing of liquids from solids has been only partially developed for compressible solids. Many aspects are in need of amplification. Herein centrifugal separation in a liquid medium is discussed. A brief presentation of centrifugal separation in a gaseous medium is also offered.

Theory

Separation by Density Difference. A single solid particle or discrete liquid drop settling under the acceleration of gravity in a continuous liquid phase accelerates until a constant terminal velocity is reached. At this point the force resulting from gravitational acceleration and the opposing force resulting from frictional drag of the surrounding medium are equal in magnitude. The terminal velocity largely determines what is commonly known as the settling velocity of the particle, or drop under free-fall, or unhindered conditions. For a small spherical particle, it is given by Stokes' law:

$$v_g = \frac{\Delta \delta d^2 g}{18 \mu}$$
(1)

where v_g = the settling velocity of a particle or drop in a gravitational field; $\Delta \delta = \delta_s - \delta_L$ = the difference between true mass density of the solid particle or liquid drop, and that of the surrounding liquid medium; d = the diameter of the solid particle or liquid drop; g = the acceleration of gravity; and μ = the absolute viscosity of the surrounding medium.

Stokes' law can be readily extended to a centrifugal field:

$$v_s = \frac{\Delta \delta d^2 \omega^2 r}{18 \mu} = v_g \left(\frac{\omega^2 r}{g} \right)$$
(2)

where v_s = the settling velocity of a particle or drop in a centrifugal field; ω = the angular velocity of the particle in the settling zone; and r = the radius at which settling velocity is determined. Analogous equations describe the terminal velocity of a light particle or drop rising in a heavier continuous medium.

The settling velocity, v_s , is relative to the continuous liquid phase where the particle or drop is suspended. If the liquid medium exhibits a motion other than the rotational velocity, ω , the vector representing the liquid-phase velocity should be combined with the settling velocity (eq. 2) to obtain a complete description of the motion of the particle (or drop).

These concepts are used to analyze separations in the bottle centrifuge, the imperforate bowl centrifuge, and the disk centrifuge. Separation by density difference in other types of centrifuges can be analyzed by analogy.

The Bottle Centrifuge. Analysis of the performance of a bottle centrifuge is based on the model shown in Figure 1. A solid or liquid particle is considered in an initial position, X, at a radius, r_i from the axis of rotation. If equation 2 is applied to this specific particle, assuming that $v_s = dr/dt$, then

$$\int_{r_i}^{r_C} \frac{dr}{r} = \int_0^t v_g \left(\frac{\omega^2}{g} \right) dt$$
(3)

where r_C = the radius of the sedimented cake, and t = the time during which the particle is subjected to centrifugal acceleration. Integration of equation 3 leads to the following:

$$\ln \frac{r_C}{r_i} = v_g \left(\frac{\omega^2}{g} \right) t$$
(4)

A radius, r , that divides the volume of supernatant into two equal parts can be defined as follows:

$$(r - r_1) = (r_C - r) \text{ or } r' = \frac{r_C + r_1}{2}$$
(5)

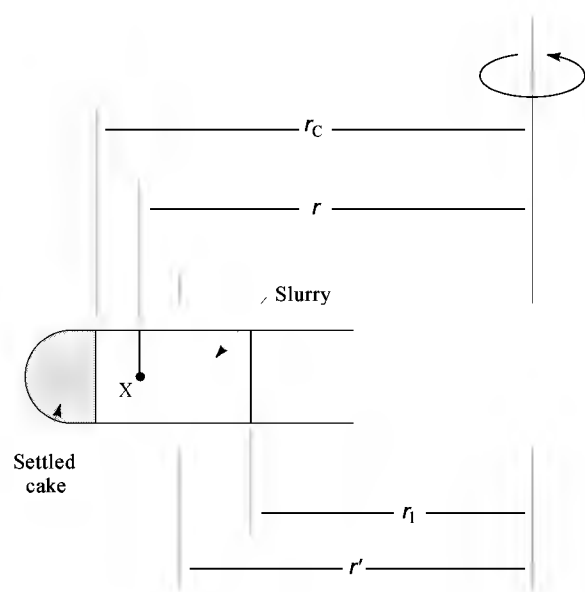


Fig. 1. Separation in a bottle centrifuge, where X is the initial position of a particle (drop). See text for definition of terms.

where r_1 is the radius of free surface of liquid. Assuming the presence of more than one particle in the feed as well as a uniform initial particle distribution, then each of the two volumes defined by r_1' initially contains the same number of particles. By making the further assumption that the particles in the suspension are identical, a settling time, t , is chosen so that those particles starting from radius r_1' all reach the cake r_c after time t . Under these conditions, one-half of the particles that were in suspension at $t = 0$ are sedimented after the time, t , has elapsed. The other half, initially located above the level defined by the radius r_1' , remain in suspension. If r is replaced (eq. 4) by r_1' , a sedimentation condition is established that is referred to as 50% cutoff. An effective capacity, Q_0 , for the bottle centrifuge is determined by the ratio between the volume, V , occupied by the slurry in the bottle, and the spinning time, t , calculated from equation 4:

$$Q_0 = \frac{V}{t} 2 v_g \left(\frac{\omega^2}{g} \frac{V}{2 \ln 2 r_c (r_c + r_1)} \right) \tag{6}$$

In equation 6, v_g characterizes the settling behavior of the solid particles or liquid drops in the suspension, whereas the second part of the right-hand side refers to speed and size of the centrifuge and is expressed by the capacity factor Σ_B . For a bottle centrifuge, it takes the following form:

$$\Sigma_B = \frac{\omega^2 V}{2 g (2 r_c (r_c + r_1))} \tag{7}$$

This capacity factor has the dimension of an area and represents the area of a static gravity settling tank having a separation performance equal to that of the rotating bottle centrifuge handling the same particles. By combining equations 6 and 7, and by eliminating volume, V , the following relation is obtained:

$$\frac{Q_0}{\Sigma_B} = \frac{2g}{(\omega^2 t)} \cdot \ln \left(\frac{2 r_c}{r_c + r_1} \right) \tag{8}$$

Equation 8 provides the basis of comparison for the performance of various bottle centrifuges having the same material, and also, under certain circumstances, of other types of sedimentation centrifuges, if geometric dissimilarities are also considered.

The capacity factor, Σ_B , defined by equation 7, is derived from a set of assumptions. An additional assumption is specific to the bottle centrifuge. Namely, a particle is considered sedimented when it reaches the surface of the cake without contacting the tube wall.

The Imperforate Bowl Centrifuge. In an imperforate bowl centrifuge the flow of the continuous liquid phase is nominally axial, except for areas immediately adjacent to the feed inlet and effluent outlet. Tubular solid-bowl basket and imperforate bowl conveyor-discharge centrifuges satisfy this definition.

The mathematical model chosen for this analysis is that of a cylinder rotating about its axis (Fig. 2). Suitable end caps are assumed. The liquid phase is introduced continuously at one end so that its angular velocity is identical everywhere with that of the cylinder. The flow is assumed to be uniform in the axial direction, forming a layer bound outwardly by the cylinder and inwardly by a free air-liquid surface. Initially the continuous liquid phase contains uniformly distributed spherical particles of a given size. The concentration of these particles is sufficiently low that their interaction during sedimentation is neglected.

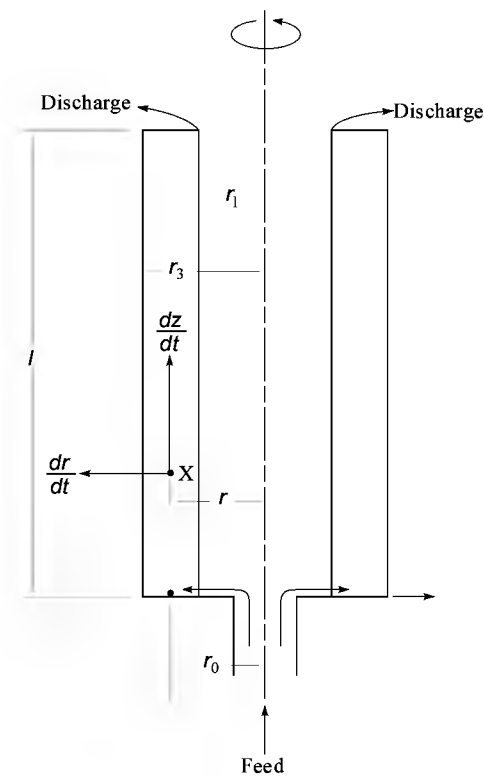


Fig. 2. Separation in a basket or tubular centrifuge, where X is the initial position. Other terms are defined in the text.

Under these circumstances, the settling motion of the particles and the axial motion of the liquid phase are combined to determine the settling trajectory of these particles. The trajectory of particles just reaching the bowl wall near the point of liquid discharge defines a minimum particle size that starts from an initial radial location and is separated in the centrifuge. A radius r is chosen to divide the liquid annulus in the bowl into two equal volumes initially containing the same number of particles. Half the particles of size d present in the suspension are separated; the other half escape. This is referred to as a 50% cutoff.

The feed rate corresponding to this condition is related to the bowl geometry r_1 , r_3 , and l , the bowl angular speed, ω ; and the Stokes' settling velocity, v_g (eq. 2).

$$Q_0 = 2v_g \left(\frac{\omega^2}{g} \right) 2\pi l \left(\frac{3}{4} r_3^2 + \frac{1}{4} r_1^2 \right) \tag{9}$$

where v_g = the Stokes' settling velocity (see eq. 1); ω = the angular velocity of the centrifuge; g = the gravitational acceleration; l = the length of settling zone; r_3 = the radius of the inside wall of the cylinder; and r_1 = the radius of the free surface of the liquid layer in the cylinder. Equation 9 can be rewritten as equations 10 and 11:

$$Q_0 = 2 v_g \Sigma_T \tag{10}$$

where

$$\Sigma_T = 2\pi l \left(\frac{\omega^2}{g} \right) \left(\frac{3}{4} r_3^2 + \frac{1}{4} r_1^2 \right) \tag{11}$$

Equation 10 estimates the flow or throughput rate, above which particles of size d are less than 50% sedimented, and below which over 50% are mostly collected. Equations 10 and 11 are also applicable to the light particles rising in a heavy phase liquid, provided that r_3 and r_1 are interchanged in equation 11.

The theoretical capacity factor, Σ_T , defined by equation 11 has the dimension of an area and can be interpreted as the area of a gravity settling tank where the separation performance is equal to the centrifuge provided that the factor v_g is the same for both. This restriction is required because the particles suspended in the continuous phase can be deaggregated and further dispersed by the vigorous shearing to which the feed is subjected during acceleration in the centrifuge. If this effect is not considered in comparison with that of the settling tank, centrifugal sedimentation might be less favorable in practice than is anticipated by theory.

Equation 11 can be further reduced to facilitate understanding of its use and application. If, instead of the radii of pond surface, r_1 , and bowl wall, r_3 , a mean radius, r_m , is introduced, equation 11 can be rewritten as follows:

$$\Sigma_T = 2\pi r_m l \left(\frac{\omega^2 r_m}{g} \right) \tag{12}$$

Equation 12 shows that Σ_T can be expressed as the product of a mean sedimentation area ($2 \pi r_m l$) and the G level ($\omega^2 r_m / g$), and therefore reflects the increased sedimentation rate expected through a defined area having centrifugal acceleration instead of gravity.

Disk Centrifuge. The separation of particles inside a disk stack is illustrated in Figure 3. The continuous liquid phase, containing solid or liquid particles to be separated, flows from the outside of the disk stack, having radius r_2 , to the inside discharge opening, having radius r_1 . Assuming that the liquid phase is evenly divided between the spaces formed by the disks, the flow in each disk space is Q_0/n , where Q_0 is the total flow through the entire disk stack and n is the number of spaces (disks). The flow of the continuous liquid phase is also assumed to be in a radial plane and parallel to the surfaces of the disks, or as having the same angular velocity as the stack.

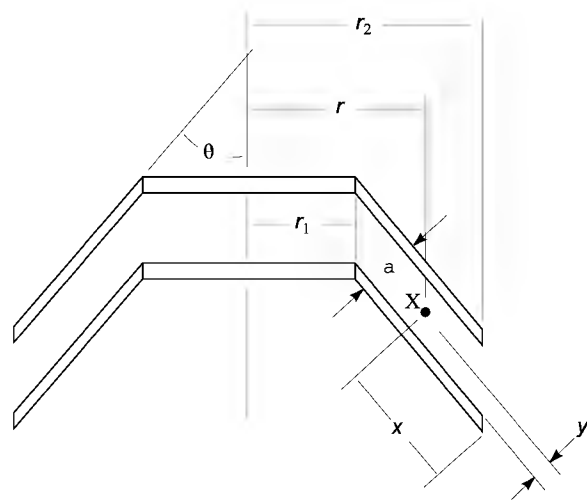


Fig. 3. Separation in a disk centrifuge, where X is the initial position of the particle. Other terms are defined in the text.

Here again an equation is established (2) to describe the trajectory of a particle under the combined effect of the liquid transport velocity acting in the x -direction and the centrifugal settling velocity in the y -direction. Equation 13 determines the minimum particle size which originates from a position on the outer radius, r_2 , and the midpoint of the space, a , between two adjacent disks, and just reaches the upper disk at the inner radius, r_1 . Particles of this size initially located above the midpoint of space a are all collected on the underside of the upper disk; those particles initially located below the midpoint escape capture. This condition defines the throughput, Q_0 , for which a 50% recovery of the entering particles is achieved. That is,

$$Q_0 = 2v_g \left(\frac{2\pi n \omega^2}{3g} \cot \theta (r_2^3 - r_1^3) \right) \tag{13}$$

where v_g = the Stokes' settling velocity (see eq. 1); n = the number of spaces between disks in a stack; ω = the angular velocity of the centrifuge; τ = half the included angle of the disks; r_2 = the outer radius of the disks; and r_1 = the inner radius of the disks (see Fig. 3). If v_g , which describes the settling characteristics of the particles, is separated from parameters relating to the geometry and rotational speed of the disk stack, then,

$$Q_0 = 2v_g \Sigma_D \tag{14}$$

where

$$\Sigma_D = \frac{2\pi n \omega^2}{3g} (\cot \theta (r_2^3 - r_1^3)) \tag{15}$$

Again, Σ_D corresponds to the area of a gravity settling tank capable of the same separation performance as the disk stack defined by the parameters included in equation 15.

Separation by Drainage. The theory covering drainage in a packed bed or particles is incomplete, and requires more development for a centrifugal field. Liquid is held within the bed by various forces. Removal involves several flow mechanisms. In addition, the centrifugal acceleration changes with radius in the bed, causing changes in packing tendencies of particles and accelerating forces on the residual liquid.

There are three types of liquid content in a packed bed: (1) in a submerged bed, there is liquid filling the larger channels, pores, and interstitial spaces; (2) in a drained bed, there is liquid held by capillary action and surface tension at points of particle contact, or near-contact, as well as a zone saturated with liquid corresponding to a capillary height in the bed at the liquid discharge face of the cake; and (3) essentially undrainable liquid exists within the body of each particle or in fine, deep pores without free access to the surface except perhaps by diffusion or compaction.

The last type of liquid can be removed by evaporation (qv) or displacement by another liquid but cannot be removed by simple flow in either a gravitational or centrifugal field. There is no sharp distinction between the first two types. The rate and extent of liquid removal from a submerged bed during drainage depends on the physical characteristics of the components, the force of the centrifugal field, and the time of exposure to the field. The residual liquid content of a drained cake consists largely of the capillary and irremovable types at the time of discharge from the separation equipment.

During cake formation and drainage the liquid moves into and through the bed in three different ways. During cake deposition, a continuous head of liquid ranging in composition from feed to clarified supernate may exist over the deposited cake. After feed is stopped, a layer of essentially clarified liquid may still exist over slow-draining cakes. Wash liquor, if it is used, may also create a liquid layer over the deposited cake. Drainage under these conditions requires continuous flow through the cake. The interstitial spaces are assumed to be full. When the free liquid layer no longer exists above the cake, the free liquid surface moves through the cake to an equilibrium position at the capillary height, leaving behind the larger voids filled with gas or vapor. Then, after bulk drainage of the larger voids, liquid still exists in the cake's upper zone in a film covering the surfaces of the solids and in partially filled voids having very restricted outlets. In time, some of this liquid flows as a film to the continuous liquid layer at the capillary height.

If a cake is sufficiently impermeable to permit the buildup of a feed or wash liquid head, flow through the cake approaches steady-state conditions except for changes in compaction or cake thickness as more feed is added, or in the liquid head if the drainage rate differs from the liquid rate addition. An equation for full-pore flow in a centrifugal field has been developed. The hypotheses set forth and for the most part proved are as follows: flow radiates out from the rotation axis and the effect of the gravitational field is negligible; voids at all points are filled with liquid that moves in laminar flow through the cake; kinetic energy changes of the liquid in the cake may be neglected, ie, the filter medium is sufficiently permeable so that it does not run full of liquid; and ambient pressure exists at the outer face of the cake. Essentially incompressible solids produce very similar cake permeabilities in a vacuum, under pressure, and during centrifugal filtration, although significant local variations in permeability may occur because of irregularities in the feed and its distribution. When pores are filled, the flow rate of the supernatant liquid through the cake (4,5) is as follows:

$$Q = \left(\frac{\pi K \omega^2 h}{\mu_L g} \right) \left(\frac{r_m^2 - r_L^2}{\ln (r_m / r_C)} \right) \tag{16}$$

where K is the permeability; h is the basket height; r_C and r_m are radii from the axis of rotation to the inner and outer faces of cake, respectively; ω is the angular velocity; and r_L is the radius from the axis to the inner face of the liquid layer. Comparison to the usual term α for cake resistance in pressure filtration shows that $K = \delta_L g / \alpha$. The functions related to cake thickness, $(r_m - r_C)$; liquid layer thickness, $(r_m^2 - r_L^2)$; angular velocity, ω^2 ; and viscosity, μ_L , have been verified by experiment. The exponent on the angular velocity varies for materials that exhibit cake compression as speed is increased, but

compression effects are minor on starch, chalk, and kieselguhr, ie, loose or porous diatomite (qv). In practice, the exponent of ω can probably be assumed to have a value of two and experimental variations in permeability can be absorbed by changes in the permeability coefficient. The real value of equation 16 lies in its use for estimating the effect of changes in operating variables for a material where the characteristics are already known from experimental data or plant operation.

Low permeability cakes draining under the conditions of equation 16 are usually handled in perforate basket centrifuges that have relatively long (20 min to several hours) cycles. Following elimination of the supernatant liquid, unsteady-state drainage of the cake may often be neglected. These slow-draining cakes often support such a large capillary height, eg, 90–99% of void volume for chalk (5), that little additional dewatering (qv) is obtained after completion of free liquid drainage. Solids of larger particle size, or freer drainage characteristics, are handled in automatic perforate baskets or continuous screen centrifuges. Low final moistures are usually achieved. For cakes of high permeability, the period when a liquid layer exists above the cake is either short or nonexistent. The time cycle depends chiefly on the rate of film drainage at the completion of bulk liquid flow. Under these conditions film drainage and permanent residual moisture are most important.

The quantity of undrainable residual moisture cannot be predicted without the benefit of experimental data. Equation 17 (6) indicates the important parameters where the exponents were determined using limited experimentation. Introducing the approximation that $s\delta_s$ is proportional to $1/\bar{d}$, where s is the specific surface area per weight of solid, the modified equation for undrainable liquid becomes

$$S^\infty = k \left(\frac{1 - \epsilon}{\epsilon} \right) \left(\frac{1}{\bar{d}^2 G} \right)^{1/4} \left(\frac{\sigma \cos \Psi}{\delta_L} \right)^{1/4}$$

(17)

where S^∞ is the fraction of void volume occupied by liquid after infinite drainage time, k is an experimental coefficient, ϵ is the void fraction, $\bar{d}$ is the mean particle diameter, G is $\omega^2 r/g$, σ is the surface tension, and Ψ is the wetting angle. Appreciable internal porosity of the particles can badly distort an experimental value of S .

For cakes of high permeability, the capillary drain height may be an insignificant fraction of cake thickness, and film drainage becomes the controlling factor in a centrifugal field (7). Under unsteady-state conditions, equation 18 represents the drainable liquid left in the cake as a function of the centrifugal filtration parameters:

$$S = S^\infty + \frac{3}{\pi} \frac{s'}{\epsilon} \omega \left(\frac{\mu_2}{\delta_L} \right) \left(\frac{h}{2 r_m h} \right) \left(\frac{1}{t'} \right)^{1/2}$$

(18)

where S is the fraction of void volume occupied by liquid at time t , s' is surface area/volume of cake, h is cake thickness, t' is time at which free liquid surface enters the cake, and μ_2 is the viscosity of the surrounding medium.

It is difficult to obtain void volume data for a cake under drainage conditions in a centrifuge. Prediction of these values from filter cake data is uncertain because the compressive force in centrifugation increases with radius throughout the cake depth, and the effective mass of a particle, proportional to $(\delta_s - \delta_L)$ when the cake is submerged, becomes essentially δ_s after bulk drainage is completed. For engineering purposes, it is simpler to approximate the ratio of the volume of liquid to the volume of solid. Assuming that $G \approx \omega^2 (2r_m - h)/g$, and $s' \approx 6(1 - \epsilon)/\bar{d}$ as for spheres, equation 18 becomes

$$q = q^\infty + \left(\frac{k'}{d} \right) \left(\frac{\mu_L}{\delta_2} \right)^{1/2} \left(\frac{h}{Gt} \right)^{1/2} \cdot 100$$

(19)

where q is the ratio of the volume of liquid to the volume of solid. This measurement is readily obtained from the volume percentage of liquid in cake and is also easily converted by a density ratio to a weight ratio of liquid to solid. The value for q^∞ , the ratio of the liquid volume to the volume of solid at infinite time, may also be applied when known.

Reasonable experimental agreement was obtained (7) using the exponents given in equation 19 for relatively slow-draining chalk and kieselguhr. Figure 4 (8) shows the effect of drain time after the disappearance of a free liquid head above the cake. The sharp break probably indicates completion of bulk drainage and start of drainage by film flow only. Drainage before the break is rapid and proportional to $t^{-1/2}$. After the break in the curve, the value $t^{-0.3}$ indicates that film drainage becomes controlling if low residual moisture is required. This exponent, 0.3, is appreciably lower than the 0.5 in theory but is close to 0.25, the value previously obtained (9). Considerable data indicate the validity of $G^{-1/2}$; limited data corroborate the theoretical exponent of $1/2$ for kinematic viscosity.

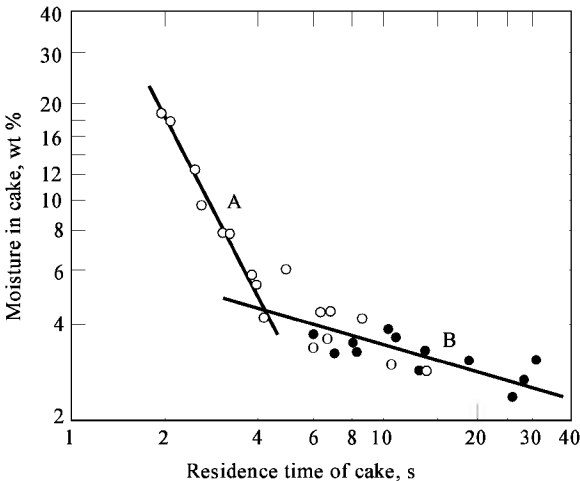


Fig. 4. Drainage of salt crystals in a cylindrical screen pusher-discharge centrifuge (8), where the cake thickness is 3.3 cm, the centrifugal field = 320 G , and the crystals 14 wt % < 250 μm . (●) Represents moisture in the discharge cake, and (○) moisture in the cake by material balance with drainage flows; line A has slope = 1.9 and B, 0.3.

Experimental exponents for cake thickness vary from 0.5 to as much as 3.0. The theoretical value of $1/2$ may be approached only by incompressible cakes of a narrow range of sizes. The proper and characteristic value for the mean particle size, $\bar{d}$, is difficult to ascertain. In practice, the most finely divided particles, eg, 10–15 wt % of solids, almost wholly determine the liquid content of a cake, regardless of the rest of the size distribution. It seems reasonable to use a $\bar{d}$ closely related to liquid content, eg, the 10% point on a cumulative weight-distribution curve.

Σ -Concept and Its Application. The assumptions and conditions for deriving equations 7, 11, and 15 impose limitations on the application of the Σ -concept and fall into two groups. The first concerns the particulate material. Particles (or drops) are assumed to be spherical in shape and uniform

in size. These should not deaggregate, deflocculate, coalesce, or flocculate during passage through the zone in which separation occurs. Initially, particles are evenly distributed in the continuous liquid phase, where their concentration is low enough for them to settle as individual particles without interaction. The settling velocity, v_s , of the particles is such that the Reynolds number does not exceed 1.0, thus ensuring that the deviation from Stokes' law does not exceed about 10%. The settling velocity, v_g , in a gravity field, or v_c in a centrifugal field, is theoretically never reached because the accelerating time required for a particle to reach its terminal velocity is infinite. However, a particle of up to 100 μm approaches 90% of its terminal velocity in milliseconds. The time available for the particles to settle in a centrifuge is enough for each individual particle to be almost at its theoretical settling velocity, despite variation in G .

The second group of assumptions and conditions concerns flow conditions. Flow is assumed to be streamlined. Fresh feed is introduced uniformly into the full space available for its flow. In an imperforate bowl centrifuge, this condition requires that the continuous liquid phase immediately occupy the full liquid layer thickness between the free surface and the inside radius of the cylindrical wall. In a disk centrifuge, the continuous phase is assumed to divide evenly between all the disk spaces axially as well as circumferentially. In the disk stack, flow lines of the continuous phase are directed radially everywhere. In any radial plane, the velocity profile normal to the disk surface is a symmetrical plug flow. In any imperforate bowl centrifuge, the continuous phase rotates everywhere at the same angular velocity as the bowl, ie, there is no forward or back swirl. The displacement of the flow pattern of the continuous phase by the layer of deposited material may be neglected. Remixing at the interface of the separated material is negligible. Finally, the detrimental effect resulting from heavy separated material crossing the fresh feed stream outside of the disk stack may be neglected.

Few of the assumed conditions are fully satisfied in practice. The last three items relate to potential interference between separated phases. Such interference can occur and leads to poor sedimentation performance if an excessive volume of the sedimented phase is retained in the centrifuge.

Excessive volume of solids may be retained in the bowl of conveyor centrifuges if (1) the conveyor volumetric displacement is not sufficient to handle the sedimentation rate of solids; (2) the sedimented solids cannot be successfully conveyed and discharged over the solids port until a sufficient layer has been built up inside the bowl; and (3) solids do not easily slide outwardly on the underside of the disk of a disk centrifuge.

In the case of the nozzle disk centrifuge, the flow of the solids phase through the discharge nozzles may be so restricted that an excessive layer can accumulate inside the bowl shell. When this layer reaches the zone utilized by the fresh feed stream entering the disk stack, reentrainment of the sedimented solids by the fresh feed may lead to poor sedimentation performance.

The sedimentation phenomenon that the Σ -concept attempts to describe quantitatively is only part of the total task that the centrifuge has to accomplish. Thus, attempts to predict separation performance solely on the basis of Σ -concepts have sometimes given disappointing results.

Nevertheless, the Σ -concept is a valuable tool, allowing in theory a comparison between geometrically and hydrodynamically similar centrifuges operating on the same feed material. Equations 7, 11, and 15 show that the sedimentation performance of any two similar centrifuges having the same feed suspension is the same if the quantity Q_0/Σ is the same for each. In practice, an efficiency factor, e , is often introduced to extend the use of Σ so as to compare dissimilar centrifuges. This factor takes into consideration differences in feeding, discharging, flow, turbulence, and remixing that exist in different types of centrifuges operating on the same feed material. The flow rate, Q_0 , of a No. 2 centrifuge can thus be compared to the rate, Q_0 , of a No. 1 centrifuge operating on the same feed. For equal sedimentation performance,

$$\frac{Q_{0_2}}{\Sigma_2 e_2} = \frac{Q_{0_1}}{\Sigma_1 e_1} \quad \text{or} \quad Q_{0_2} = Q_{0_1} \left(\frac{e_2}{e_1} \right) \left(\frac{\Sigma_2}{\Sigma_1} \right)$$

(20)

If the two centrifuges are geometrically and hydrodynamically similar, then $e_1 = e_2$ and equation 20 can be simplified to

$$Q_{0_2} = Q_{0_1} \frac{\Sigma_2}{\Sigma_1}$$

(21)

The relation of equation 21 for similar centrifuges requires identical sedimentation performance characteristics when operating on the same material.

The Σ -concept permits scale-up between similar centrifuges solely on the basis of sedimentation performance. Other criteria and limitations, however, should also be investigated. Scale-up analysis for a specified solids concentration, for instance, requires knowledge of solids residence time, permissible accumulation of solids in the bowl, G level, solids conveyability, flowability, compressibility, limitations of torque, and solids loading. Extrapolation of data from one size centrifuge to another calls for the application of specific scale-up mechanisms for the particular type of centrifuge and performance requirement.

Other Sedimentation Scale-Up Equations. Some centrifuge suppliers use an area-equivalent, A_e , description instead of Σ ; others use KQ or Lf_2 values. All of these are in units of area. For a disk centrifuge,

$$\Sigma_D = \frac{2\pi n \omega^2}{3g} \cot \theta (r_2^3 - r_1^3)$$

(15)

$$KQ = \frac{2\pi n \omega^{1.5}}{3g} \cot \theta (r_2^{2.75} - r_1^{2.75})$$

(22)

$$Lf_2 = \frac{2\pi n \omega^2}{3g} \cot \theta \left(r_2^3 - \frac{r_1}{r_2} \left(\frac{r_1}{r_2} \right)^2 \left(\frac{r_1}{r_2} \right)^3 \right)$$

(23)

For an imperforate tubular centrifuge,

$$A_{e_{3/4}} = \frac{2\pi l \omega^2}{g} (0.75 r_{0.75}^2)$$

(24)

$$\Sigma = \frac{2\pi l \omega^2}{g} (0.75 r_3^2 + 0.25 r_1^2)$$

(25)

All of these equations work in the scale-up of geometrically similar centrifuges. The KQ reduces the effect of rotational speed from ω^2 to $\omega^{1.5}$ and the disk radius from r^3 to $r^{2.75}$, based on empirical experience. The Lf_2 uses the projected cylinder area of the disks at their mean radius (see eq. 12 of Σ_T for imperforate bowls). $A_{e_{3/4}}$ is the cylindrical area at 3/4 of the bowl wall radius.

When testing a new material for centrifugal separation, a bottle centrifuge is usually used to obtain the general G range needed, and to choose the centrifuge type and size. To estimate size, the Q_0/Σ_B must be determined for the bottle centrifuge (eq. 7) and then used in equation 21 to determine the Σ value of the centrifuge to be used. The efficiency factor assumed for disk and imperforate centrifuge has been found to be $\sim e = 0.5$, compared to the bottle centrifuge.

Factors Influencing Centrifugal Sedimentation. The sedimentation velocity of a particle is defined by equations 1 and 2. Each of the terms therein effects separation.

Viscosity. Sedimentation rate increases with decreased viscosity, μ , and viscosity is dependent on temperature. Often mineral oils, which are highly viscous at room temperature, have a viscosity that is reduced by a factor of 10 at 70–80°C. Tar, solid at room temperature, is a low viscosity liquid at 150–200°C and can be clarified of inorganic solids at high flow rates. Even the viscosity of water changes significantly when the temperature changes between 10 and 35°C (10).

Density Difference Between Particle and Liquid. Separation cannot take place if $\Delta\delta = 0$. Some mineral oils have the same density as water at room temperature. If it is heated to 80°C, the reduction of the density of water is less than that of the mineral oil, resulting in the water becoming heavier. Therefore separation is possible. Dilution of a liquid by a solvent, eg, molasses by water or heavy oil by naphtha, results in lower density and lower viscosity of the liquid. Solvent stripping takes place at a later stage.

Particle Size. Doubling particle size, d , increases sedimentation by a factor of 4. Thus, methods to increase size are important. Additives are commonly used to flocculate many fines into an agglomerate that acts as a single large particle. Many chemical companies offer a wide range of organic and inorganic flocculating products, in dry or liquid form. Bench tests are usually required to determine best type and dose, ie, to optimize the flocculent choice. pH control, electrostatic devices, and mechanical coalascers are used to combine fine liquid drops in emulsions to produce larger particles. Special care must be exercised to pump, pipe, and feed mixtures of these easily breakable particle agglomerates, thus preventing the large particles from becoming fine before sedimentation.

Particle Shape. Whereas the Stokes' particle is assumed to be a sphere, very few real solids are actually spherical. Flat and elongated particles sediment slower than spheres. For maximum sedimentation rate, the particle should be as spherical as possible.

Particle Size Distribution. Almost every feed slurry is a mixture of fine and coarse particles. Performance depends on the frequency of distribution of particle size in the feed. Figure 5 shows that whereas all of the coarse particles having a diameter greater than some d_{lim} are separated, fewer of the very fine particles are, at any given feed rate. The size distribution frequency of particles in feed and centrate for a fine and coarse feed are quite different. More coarse particles separate out than fine ones. Classification of solids by size is often done by centrifugal sedimentation.

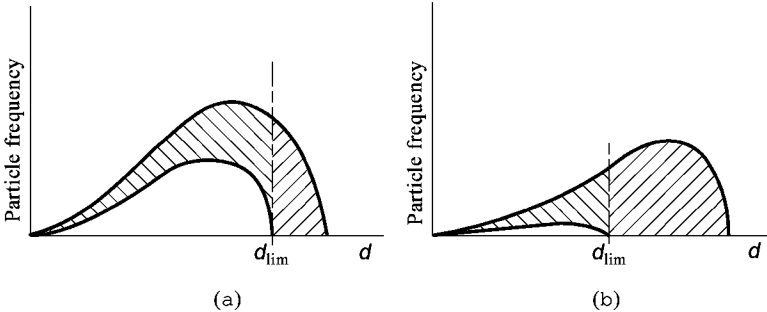


Fig. 5. Particle distribution (upper line) before and (lower line) after action of the separator where the cross-hatched areas represent the particles separated out. By definition, all particles of $d > d_{lim}$ are separated out. A number of particles having $d < d_{lim}$ are also separated. (a) Fine and (b) coarse particle dispersion (10).

Sedimentation Velocity. Velocity of sedimentation, v_s , is equal to throughput Q when divided by Σ . If the concentration in the centrate, c , is divided by that of the feed, c_0 , then c/c_0 gives the unsedimented percentage. A log-log plot of c/c_0 against the centrate rate, Q , usually results in a near-straight line for normal particle size distributions (Fig. 6) (10). In comparing the level of performance at the same percent unsedimented ratio for the same feed material and conditions, the ratio Q_1/Q_2 should equal Σ_1/Σ_2 . For materials where the size distribution is very narrow, however, the performance curve is not a straight line (Fig. 7) (10). Almost all of the particles are sedimented at rates below the break in the curve.

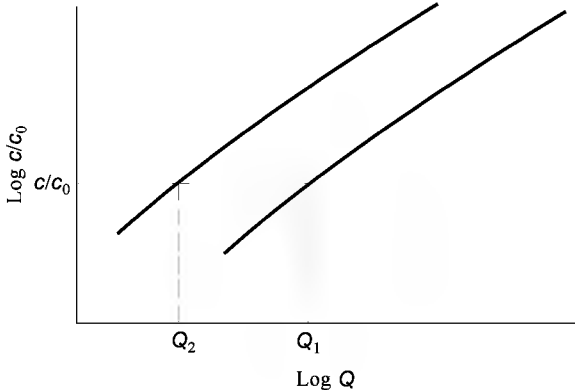


Fig. 6. Relative performance, E_{rel} , where $E_{rel} = Q_1/Q_2$.

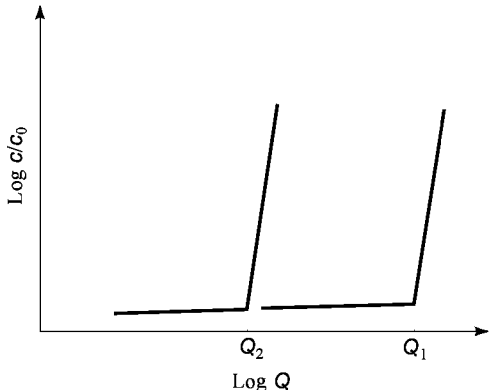


Fig. 7. Determination of E_{rel} for particle distribution within narrow limits. See Fig. 6.

Liquid–Liquid-Phase Behavior

Liquid drops, suspended in a continuous liquid medium, separate according to the same laws as solid particles. After reaching a boundary, these drops coalesce to form a second continuous phase separated from the medium by an interface that may be well- or ill-defined. The discharge of these separated layers is controlled by the presence of dams in the flow paths of the phases. The relative radii of these dams can be shown by simple hydrostatic considerations to determine the radius of the interface between the two separated layers. The radius is defined by

$$r_i^2 - r_h^2 = \frac{\delta_l}{\delta_h} (r_i^2 - r_l^2)$$
(26)

where r_i , r_h , and r_l are the radii of the interface, the liquid surface at the heavy discharge dam, and the liquid surface at the light discharge dam, respectively, and δ_h and δ_l are the densities of the heavy and light phases, respectively. Control of the interface radius, achieved by varying r_h or r_l for the desired ratio, is an important factor in liquid–liquid separation, as it determines whether the heavy or light phase is exposed to the greater clarifying effect (10).

Equation 26 is accurate only when the liquids rotate at the same angular velocity as the bowl. As the liquids move radially inward or outward these must be accelerated or decelerated as needed to maintain solid-body rotation. The radius of the interface, r_i , is also affected by the radial height of the liquid crest as it passes over the discharge dams, and these crests must be considered at higher flow rates.

Centrifuge Components

Power, Energy, and Drives. Centrifuges accomplish their function by subjecting fluids and solids to centrifugal fields produced by rotation. Electric motors are the drive device most frequently used; however, hydraulic motors, internal combustion engines, and steam or air turbines are also used. One power equation applies to all types of centrifuges and drive devices.

The total power, P_T , needed to run a centrifuge, ie, delivered by the drive device, is equal to sum of all losses:

$$P_T = P_P + P_S + P_F + P_W + P_{BD} + P_{CP}$$
(27)

where P_P , the process power, $= Q\delta\omega^2r^2$; Q is the flow rate of liquid or solids; δ = mass density; and r = discharge radius of material being discharged. Power for each liquid and the solid phase must be added to get P_P . P_S , the solids process power, $= T_C\Delta N$ for scroll decanters, where T_C = conveyor torque and ΔN = differential speed between bowl and conveyor. P_F is the friction power, ie, loss in bearings, seals, gears, belts, and fluid couplings. P_W , the windage power, $= K\mu^{0.2}\rho^{0.8}N^3D^{4.5}$ and μ = viscosity of surrounding gas; ρ = density of gas; D = rotor outside diameter; N = rpm; and K = shape constant. Increased density owing to gas pressure increases the windage power, and this may be very significant for high pressure applications. Also, many hydrocarbon gases are heavier than air, resulting in high windage power. Very high G centrifuges often operate in a vacuum to avoid excessive windage power. For constant G , the scale-up windage power $\propto D^3$; for constant rotor stress, the scale-up of windage power $\propto D^{1.5}$.

Windage power is a very important loss for large machines and must be determined. Whereas windage power can be calculated from drawing dimensions (11), it is preferable to measure the windage power for an actual rotor, and then extrapolate using the formula given for windage power for a geometrically similar (larger or smaller) size. Doubling the size of a rotor while maintaining the g level results in eight times the windage power loss.

P_{CP} is the power absorbed by the centripetal pump. The centrate kinetic energy is partially recovered by the pump, which delivers the centrate flow at a positive pressure. The added power must be supplied by the centrifuge main drive, but use of a centripetal pump avoids the need for a separate centrate pump. The power required to bring the feed to the centrifuge is supplied by a feed pump at the feed tank, not by the centrifuge driver. This power may be significant where the feed pressure required at high flow rates is high.

For scroll centrifuges having back-drives, P_{BD} is the back-drive power:

$$P_{BD} = K \frac{T_C}{R} N_{BD}$$
(28)

where T_C = conveyor torque; R = gear box ratio; and N_{BD} = back-drive speed. This power, provided by the centrifuge drive device, is absorbed by the brake as heat, or if a regenerative device is used, recovered by the electric power supply. Regenerative back-drives reduce total power consumption. A high gear box ratio results in lower back-drive power. Direct hydraulic motor conveyor drive devices get their power from an external hydraulic power supply, not from the main drive motor, and must meet the direct conveyor torque demands. The equation for direct hydraulic conveyor power is $P_{BD} = KT_C\Delta N$; however, hydraulic losses in the power supply, rotary seals, and the hydraulic motor must be added. Eddy current brakes, d-c motors, and variable-frequency a-c motors are the most common variable-speed back-drive devices.

The choice of the main drive, usually an a-c motor, must include starting specifications. Various methods are used to start centrifuges. These include mechanical, fluid, and hydraulic couplings, as well as WYE-DELTA electric motor starters and electronic motor controls. Centrifuges are high inertia rotating devices, often taking 1–12 min to accelerate to operating speed. Details of the mass moment of inertia, friction, and windage losses must be considered to specify a drive device. The inertia seen by the drive device, when comparing centrifuges operating at constant G , is proportional to D_4 . Small rotors are easy to start, but large rotors must be carefully reviewed.

Disk machines having nozzles to discharge the solids phase through small backward-pointing nozzles must have this power included in the calculations. Some thickening scroll centrifuges also use such nozzles, usually with an intermittent flow. P_N is the nozzle power:

$$P_N = Q_N\delta\omega r_N(\omega r_N - v_N \cos \phi)$$
(29)

where Q_N = the volume rate of the material discharged through the nozzles, d_N = the mass density of the material discharged through the nozzles, γ = the angular velocity of the centrifuge, r_N = the radius at which the nozzles discharge, v_N = the linear discharge velocity out of the nozzles, and ϕ = the angle measured between the direction of the nozzle and the tangent to the circle of radius, r .

The discharge velocity out of the nozzles is given by equation 30:

$$v_N = C(2p_N - \delta_N)^{1/2}$$
(30)

where C = a discharge coefficient, and p_N = the hydraulic pressure at radius r_N resulting from the rotation of the bowl. The pressure p_N is given by equation 31:

$$p_N = \frac{\delta_n\omega^2}{2}(r_N^2 - r^2)$$
(31)

where r is either the free-surface radius of the heavy liquid phase in the case of liquid–liquid and liquid–liquid–solids separation, or the free-surface radius of the liquid phase in the case of liquid–solids separation; and δ is intermediate between the mass densities of the heavy phase and that of the nozzle

stream, in the case of liquid–liquid–solids separation, or the liquid phase and that of the nozzle stream, in the case of liquid–solid separation.

The coefficient C in equation 30 is a function of two phenomena. First, the presence of viscous friction accounts for a small loss of energy. Nozzle orifice contraction results in discharge coefficients which range from 0.5 to 0.85. The coefficient falls in the upper portion of this range when the length of the nozzle is two or three times its diameter, and in the lower end of the range where the nozzle diameter is more than five times its length.

Special vortex nozzle designs which deliver lower flows using a large opening have been used to reduce plugging problems. Also, viscosity-sensitive nozzles, where flow is increased as viscosity is increased, are used to control variation in solids concentrations.

Filtering centrifuges must consider the power needed to bring the solids to a final radius, r_s . The radius used to determine the centrate power in equation 27 is the outside radius of the rotor supporting the filtering screen. Many filtering centrifuges discharge solids at a reduced bowl speed to avoid particle breakage. The main drive is often designed to recover the rotor kinetic energy during deceleration.

The energy absorbed by the liquid and solids stream, P_p , is transferred in the feed zone of the rotor or conveyor. One-half of the total energy is converted to the kinetic energy associated with the tangential velocity of the pond surface. Because the pond surface is at or near the discharge dam radius, this kinetic energy is dissipated as heat when the centrate is discharged to the stationary casing. The other half of the total energy is lost as turbulence in or near the feed zone. The intensity of this turbulence can break friable particles, or create tight emulsions of two immiscible liquids, both making the separation which follows more difficult. Reducing the pond radius reduces the total process power and particle degrading and thus improves total separation performance. Reducing the pond inside radius by 25% reduces total power by 44%.

Materials of Construction and Operational Stress. Before a centrifugal separation device is chosen, the corrosive characteristics of the liquid and solids as well as the cleaning and sanitizing solutions must be determined. A wide variety of materials may be used. Most centrifuges are austenitic stainless steels; however, many are made of ordinary steel, rubber or plastic coated steel, Monel, Hastelloy, titanium, duplex stainless steel, and others. The solvents present and of course the temperature environment must be considered in elastomers and plastics, including composites.

Once the material choice based on corrosion is made, a careful analysis of the stresses produced by rotation for the particular type of centrifuge is required, so that for the given liquid and solids specific gravities a maximum operating speed can be determined. In general, the metals used are ductile and elastic in the operating speed range. The usual limits are the ultimate strength and yield strength (0.2% offset). The stresses of the centrifuge bowl are primarily tangential stress and axial stress. These are the result of the weight of the bowl material itself, and the need to contain the rotating process solids and liquids within the bowl.

The geometry of the bowl parts is important, and for intermittently discharging centrifuges, fatigue strength must be considered. In general the tangential stress in a bowl wall is σ_T .

$$\sigma_T = \sigma_{\text{self}} + \sigma_p$$
(32)

where

$$\sigma_{\text{self}} = K \delta_M \omega^2 r^2$$
(33)

$$\sigma_p = K \delta_p \omega^2 (r^2 - r_s^2)$$
(34)

where δ_m = mass density of bowl material; r = bowl radius; δ_p = mass density of process (δ_p is usually the wet solids, but can include heavy liquid phase); r_s = inside radius of the pond; and K is a function of bowl geometry and includes stress concentrations owing to changes in section, holes, slots, fillets, etc. The total stress increases with the square of bowl speed, ω^2 , and the pond radius, and is directly proportional to the liquid and solid density.

Within the bowl, the pressure, p , developed also exerts axial separation forces. The internal pressure in a bowl is shown in equation 35:

$$p = \delta_p \omega^2 (r^2 - r_0^2) / 2$$
(35)

The axial projected area $A = \pi(r^2 - r_0^2)$ and the average pressure is $1/2 p_{\text{max}}$. The axial force $F_A = p_{\text{max}} A / 2$, where

$$F_A = \delta_p \omega^2 \pi (r^2 - r_0^2)^2 / 4$$
(36)

The stress resulting from the axial forces must be considered in analyzing the parts which resist axial separation, such as bolts, nuts, rings, etc. On scroll centrifuges the axial force owing to the axial component of the conveyor torque must also be considered.

Typically the total stress on the bowl is 50–65% self-stress, and 35–50% process stress. The axial stress is usually 5–10% of the tangential stress. When the bowl material density is low (such as titanium), or where solids are heavy (as is coal), or when the pond surface is close to the inside wall (as when using shallow pond scroll centrifuges), these proportions differ. All centrifuges must have a factor of safety so that the maximum stress at any point is less than the yield strength. The ratio of the material strength to the actual stress is the factor of safety. Factor of safety is set by the manufacturer, and sometimes, especially in Europe, by government regulation. Every centrifuge supplier sets the limits of bowl speed, temperature, and liquid and solids density. On some very high G preparative or zonal centrifuges, the number of cycles of bowl use must be recorded and the rotor retired after a given number, because fatigue determines the bowl life. Most disk and scroll centrifuges are made from forgings, centrifugal castings, or fabrications. Each manufacturer must carefully monitor the strength and ductile properties of the basic bowl material to ensure that the required factor of safety is maintained. A rotor failure during operation is very serious, and can not only destroy the centrifuge, but also damage nearby equipment and injure operators.

Many of the centrifugal separation applications are abrasive and erosive to centrifuge parts because of the high relative velocity or high contact pressure between the particles and part. There are many cases of erosion and corrosion wear. Areas of wear must be protected using materials which resist both mechanical and chemical attack. In general the main areas of abrasive wear are the feed zone surfaces, where nonrotating process slurries are accelerated to speed; the tips and faces of the conveyor flights on scroll centrifuges; solids discharge openings; screens, where solids slide across the screen; and stationary collection surfaces that receive the impact of solids being discharged from the bowl. Many materials are used. Examples are sintered tungsten carbide bound with cobalt or nickel; sintered aluminum oxide ceramic; sprayed and fused or weld-deposited hard-surfacing alloys; and elastomeric coatings or inserts. There is great variation in the design and application of these materials which effect the actual working life in service and the cost to rebuild the worn areas or replace the worn insert.

Maintenance operating costs are lower if replaceable wear protection is used, compared to the requirement to rebuild (and thus rebalance) the bowl materials. In very severe applications, such as coal (qv) or coal refuse dewatering that use screens, wedge wire screens made from ceramic or carbide materials are required. The use of carbide or ceramic flight-tip protection on continuous scroll centrifuges has permitted economic use in applications such as dewatering (qv) tar sands as well as coal and coal refuse, and dewatering and thickening mixed primary and secondary sewage sludges. Continuous nozzle discharge disk and scroll centrifuges would not be feasible without replaceable ceramic or carbide nozzle inserts. The use of relatively soft abrasion-resistant elastomers, such as urethanes, has been successfully applied in the solids receiver housing and the feed zone targets of scroll centrifuges. Hard chrome plating has been used on the first few disks of disk centrifuges. Stellite or carbide inserts are often used at the solids discharge opening.

Noise. Centrifuges, as do any rotating equipment, create noise. When the motion of air or gas entrained by a rotating bowl shell is deflected or otherwise disturbed, its energy is transferred to the environment through the casing or chutes. This mechanism suggests that the noise level created by the bowl is related to the surface linear speed of the rotor. High linear speed is important for maintaining separating capacity, so the noise level should be reduced without reducing the rotational speed, if possible.

In addition to surface speed, rotor imbalance, surface irregularities, clearance between covers and rotor, resonance in the supporting structure, conditions of installation, and particularly the drive motor contribute to centrifuge noise. An inadequate supporting platform can amplify centrifuge vibration. Open piping and venting and discharge connections allow noise generated inside the unit to escape. Discharge connections should be tight, yet flexible enough to prevent transfer of vibrational energy to plant piping. Size, spacing, materials of construction, and other properties of the centrifuge room also affect noise level. Sound-absorbing materials or enclosures should be provided when other means are inadequate.

The electric motors are often the noisiest component of the centrifuge assembly. Most standard motors in the 75–250 kW range develop noise levels of 85 dbA (weighted sound pressure level using filter A, per the ANSI standard). A quiet motor can reduce this level by 5 dbA and should be used whenever noise is of concern.

Equipment. Centrifugation equipment that separates by density difference is available in a variety of sizes and types and can be categorized by capacity range and the theoretical settling velocities of the particles normally handled. Centrifuges that separate by filtration produce drained solids and can be categorized by final moisture, drainage time, G , and physical characteristics of the system, such as particle size and liquid viscosity.

For optimum results, a combination of several types of equipment may be used, eg, a gravity separator for oil recovery from sludge at a petroleum refinery. The sludge, an aqueous suspension of 1–5% oil and 5–30% solids below 50 μm , is screened to remove trash, and degrittied in a cyclone to eliminate the coarse solids that would cause excessive abrasion. An imperforate bowl conveyor-discharge centrifuge then removes 60–70% of the solids in oil-free condition. The resulting oil-in-water emulsion, stabilized by residual fine silt, is passed through a 0.25-mm (60-mesh) screen, and sent to a disk centrifuge that discharges an oil stream at 0.5–2% bottom sediment and water, and an oil-free peripheral nozzle discharge containing the remaining solids in water.

Equipment Materials and Abrasion Resistance. Stainless steel, especially Type 316, is the construction material of choice and can resist a variety of corrosive conditions and temperatures. Carbon steels are occasionally used. Rusting may, however, cause time-consuming maintenance and can damage mating locating surfaces, which increases the vibration and noise level. Titanium, Hastelloy, or high nickel alloys are used in special instances, at a considerable increase in capital cost.

Abrasion, a serious problem in some applications, requires the addition of hard-surfacing materials to points exposed to abrasive wear (12). The severity of wear depends on the nature, size, hardness, and shape of particles as well as the frequency of contact, the force exerted against the wearing parts, and solids loading as related to feed rate and solids concentration.

A wide range of abrasion-resistant materials is available. Nickel–chrome–boron and cobalt–chrome–tungstenhard-surfacing alloys have been used for many years. Composite coatings of nickel-base alloys containing crushed tungsten carbide particles, applied by flame spraying and fusing, are also used. Solid tungsten carbide, pressed to shape and sintered at high temperatures, provides the best protection. Tungsten carbide plates brazed or bonded to stainless steel supports, which in turn can be easily welded to portions of a centrifuge such as conveyor flights, have been very successful. Ceramics have been used where minor impact and abrasive particle pressures are involved.

Centrifuges

Manufacturers of sedimentation and filtration centrifuges can be found in Reference 13.

Sedimentation Equipment. Centrifugal sedimentation equipment is usually characterized by limiting flow rates and theoretical settling capabilities. Feed rates in industrial applications may be dictated by liquid handling capacities, separating capacities, or physical characteristics of the solids. Sedimentation equipment performance is illustrated in Figure 8 on the basis of nominal clarified effluent flow rates and the applicable $Q_0 \Sigma$ values. The latter are equivalent to twice the theoretical gravity settling velocities. In liquid–solid separations, the effluent rate represents the clarified stream of the liquid medium and does not include the volume of solids discharged or the volume of medium discharged with the solids. The effluent rate of liquid–liquid separation refers to the clarified, heavy, or light continuous phase that usually occupies the greater volume within the separating equipment. The flow range for a particular piece of equipment does not represent its absolute limitations, but the normal flows for good clarification in standard applications. Similarly, large particles can always be sedimented.

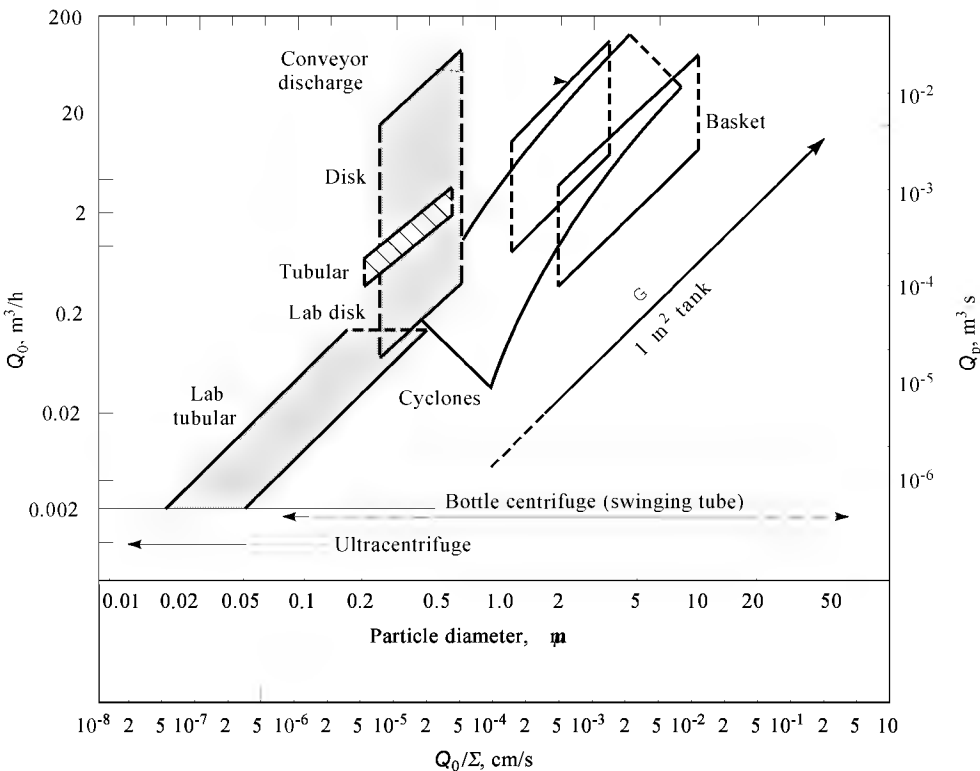


Fig. 8. Sedimentation equipment performance where the particles have a $\Delta\delta$ value of 1.0 g cm^{-3} and a viscosity, μ , value of 1 mPa s ($= \text{cP}$). The value of $Q_0 \Sigma$ is twice the settling velocity at $G = 1$, and Q_0 overflow discharge rate in measurements given.

As an additional guide, the $Q_0 \Sigma$ values are correlated with the equivalent spherical particle diameter by Stokes' law, as in equation 1. A density difference $\Delta\delta$ of 1.0 g cm^{-3} and a viscosity of 1 mPa s ($= \text{cP}$) are assumed, thus conversion to other physical characteristics of the system requires that the particle size scale be adjusted to equate a particle of $1.0 \text{ }\mu\text{m}$ diameter to its $Q_0 \Sigma$ in cm s , according to the relationship $Q_0 \Sigma = 10^{-7} \times 1.09 \Delta\delta \mu$, for $\Delta\delta$ in g cm^{-3} , and viscosity μ in Pa s . For interpretation of the particle sizes, the scale refers to the 50% cutoff particle size, and under actual centrifugation conditions the value of Σ , determined from Figure 8, must be increased by efficiency factors to give the theoretical value of Σ .

Figure 8 serves as a guide to the types of equipment that can handle a given separation. Other characteristics further narrow selection. For example, for the separation at $Q_0 = 3.5 \times 10^{-3} \text{ m}^3/\text{s}$ (50 gpm) of kaolin clay solids from an aqueous suspension, where the particle density is 2.55 g/cm^3 and the size ranges from 0.25 to $30 \text{ }\mu\text{m}$ with 55% $<2 \text{ }\mu\text{m}$, the $1.0\text{-}\mu\text{m}$ point on the particle size scale would be equivalent to $Q_0/\Sigma = 1.69 \times 10^{-4} \text{ cm}^2/\text{s}$. Assuming that a high recovery is desired, a disk centrifuge is required, and recovery of most particles greater than $0.4 \text{ }\mu\text{m}$ is satisfactory, the Q_0/Σ equivalent to $0.4 \text{ }\mu\text{m}$ on the adjusted scale is about $2.3 \times 10^{-5} \text{ cm}^2/\text{s}$. Using a disk machine efficiency of 40%, the centrifuge needed would have a Σ value of

$$\Sigma = \frac{3.15 \times 10^{-3} \times 10^6}{0.40 \times 2.3 \times 10^{-5}} = 34.3 \times 10^7 \text{ cm}^2$$

Because clay tends to pack hard, only the continuous nozzle discharge bowl would be satisfactory. Intermittently discharging disk centrifuges could not be used.

If classification of solids were desired, several other types of centrifuges could be used as well, assuming that only particles over about $2 \text{ }\mu\text{m}$ were to be removed from the suspension and that the oversize stream should be highly concentrated for disposal. Figure 8 shows that a conveyor-discharge or basket centrifuge or standard cyclone could theoretically be used. Because cyclones cannot concentrate the oversize as much as the centrifuges, these are less satisfactory for this example. If the feed concentration were low, eg, less than a few percent solids, a basket centrifuge would be used with intermittent discharge of solids. A conveyor-discharge centrifuge gives almost as good a concentration of oversize as the basket and is a more efficient classifier. The conveyor-discharge centrifuge would thus be the better choice and is actually used in the kaolin industry.

In general, solids-retaining batch and batch automatic machines are limited to low feed concentrations to minimize the time required to unload the solids. Continuous disk centrifuges can have higher feed concentration. The limit is the underflow concentration. Conveyor discharge centrifuges can handle high feed concentration and are limited only by the volume of solids displacement, or torque capacity.

Using flocculent to create aggregated solids, varying degrees of deflocculation of the solid particles may occur during acceleration in the centrifugal field. An additional problem is the removal of the resulting soft, slimy cakes under centrifugal force. Scroll centrifuges have been successfully used to discharge soft, slippery solids continuously. Specially modified, automated basket centrifuges can handle a broad range of soft sludges, often without polymer addition. Disk centrifuges are particularly well-suited for clarification of streams containing solids such as aluminum hydroxide or waste-activated sludge.

The disk centrifuge having high capacity and G level is normally used for separating a liquid-liquid mixture or for clarification of such a mixture containing fine solids (14). Specially modified conveyor centrifuges are also used for three-phase separations. Settling velocity is a criterion of selection, but the actual separating of emulsified liquid-liquid mixtures may not strictly follow a settling theory. To break an emulsion, a threshold level of centrifugal force may have to be exceeded. In addition, drainage of liquid from the continuous-phase film of the emulsion has a time factor. Centrifugal force and time cannot be calculated interchangeably as Σ -theory would indicate. In centrifugation equipment, coalescence of the dispersed liquid occurs coincidentally with its separation because there is neither time nor space for appreciable interfacial retention of unbroken emulsion.

Batch equipment, such as the bottle centrifuge or ultracentrifuge, does not have a real throughput capacity. By increasing the time of operation, according to Figure 8, the smallest particle size of solids usually sedimented in the bottle centrifuge may be $0.1 \text{ }\mu\text{m}$, at which size Brownian movement controls. In the ultracentrifuge, separation can be achieved down to molecular size, perhaps $0.005 \text{ }\mu\text{m}$, where Brownian movement is controlling. There is clearly no limitation to the larger particles that may be settled in bottle centrifuges so that an arbitrary upper limit is indicated for practical minimum conditions of 1000 rpm and 10 s. Similarly, for the ultracentrifuge the upper limit for Q_0/Σ was estimated for minimum conditions of one hour and 5000 rpm.

Centrifugal Sedimentation Equipment. Commercial sedimentation centrifuges are characterized principally by how solids are discharged, and the general dryness of these solids. There are batch and automatic batch solid bowl machines which collect the solids at the bowl wall. Solids are removed very dry. Almost any solid is collectable, even those that are very soft and compressible.

Disk-type solid bowl machines are batch, batch automatic, and continuous. The solids are removed in many different ways, but are usually wet. Scroll centrifuges discharge solids continuously and usually drier than disk and imperforate batch types. Generally disk centrifuges have the highest values of Σ or KQ for a given size and therefore the best ability to collect fine particles at a high rate.

Bottle Centrifuge. A bottle centrifuge is designed to handle small batches of material for laboratory separations, testing, and control. The basic structure is usually a motor-driven vertical spindle supporting various heads or rotors. A surrounding cover reduces windage, facilitates temperature control, and provides a safety shield. Accessories include timer, tachometer, and manual or automatic braking. Bench-top bottle centrifuges operate at 500–5000 rpm, producing centrifugal fields up to 3000 G in the lower speed range, and operate up to 20,000 rpm with 34,000 G in the high speed units. Larger models operate up to 6000 rpm and develop 8000 G , using special attachments that permit 40,000 G . These models may also be equipped with automatic temperature control down to -10°C and other programmable controls to manage the cycle.

There are three types of rotors: swinging bucket, fixed-angle head, or small perforate or imperforate baskets for larger quantities of material. In the swinging bucket type, the bottles are vertical at rest but swing to a horizontal radial position during acceleration so that solids are deposited in a pellet at the bottom of the tube. Although sedimenting particles must travel up to the full depth of the liquid layer, which requires appreciable time, the long path of travel and the perpendicularity of the sedimenting boundary to the axis of the tube are distinct advantages in effecting fractional sedimentation. Heads carrying fixed tubes at a $35\text{--}50^\circ$ angle reduce centrifuging time because the maximum distance traveled by a particle is the secant of the tube angle times the diameter of the tube. Particles strike the wall and slide down the tube to collect near the bottom, but the angle makes it difficult to measure relative volumes of supernatant liquid and sedimented solids. Rotors carry 2 to 16 metal containers having tubes and bottles of various sizes and shapes. Containers range in capacity from capillaries for microanalysis to a 1-L maximum, limiting the batch capacity of this type of centrifuge to 4 L. Although glass bottles and tubes are generally used, plastic and metal containers are available for high speed operation or corrosive liquids. Tubes are usually cylindrical, tapered, and graduated; special shapes for analytical work are available, including pear-shaped tubes having capillary tips for measuring small quantities of solids.

The bottle centrifuge is primarily used in the laboratory to separate small quantities of material. It is also used for standard analyses including many ASTM methods and in preliminary testing for scale-up to commercial centrifuges. The Σ_B value for free-swinging tubes is determined by equation 7 and the Q_0/Σ_B value by equation 8; Q_0/Σ_B data can be prepared by bottle centrifuge (5,15). The bottle centrifuge has been used to study waste treatment sludges for estimation of cake concentrations and feasibility of handling in a conveyor centrifuge (16). Because compaction of solids is largely a function of the centrifugal field force and the exposure time, the bottle centrifuge can also be used to study these parameters, although not always in the range applicable to industrial equipment. Similarly, drainage of packed solids can be studied by using tubes with fretted glass or perforated metal bottoms. Closed containers should be used to prevent drying by windage.

Specialty rotors permit ordinary bottle centrifuges to achieve some of the results previously considered possible only in ultracentrifuges. A modified zonal rotor, shown in Figure 9, permits collection of sediment using continuous addition of feed and discharge of centrate.

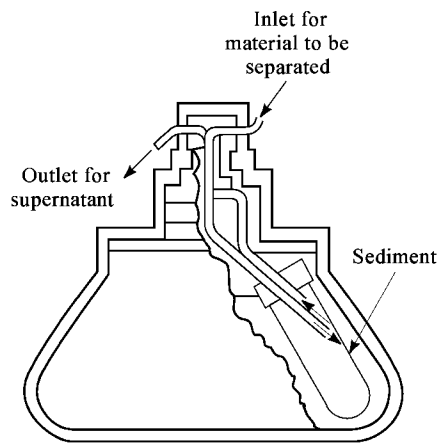


Fig. 9. Tube-type continuous-flow rotor. Courtesy of Sorvall.

Preparation Ultracentrifuge. Preparation ultracentrifuges are suitable for a range of applications, such as processing quantities of subcellular particles, viruses, and proteins (qv). Many design variations are available and only the common features are considered here. Preparation ultracentrifuges range in operating speed from 20,000 rpm, generating about 40,000 *G*, to 75,000 rpm and about 500,000 *G*. The rotor is surrounded by a high strength cylindrical casing and underdriven by an electric motor. To avoid overheating of the rotor by air friction at these speeds, the pressure in the casing is reduced to about 0.13 Pa (1 μ m Hg). Sensors (qv) monitor the temperature and a cooling system controls the temperature in the range of $\pm 1^{\circ}\text{C}$. Electronic controls maintain the rotor speed within a required narrow range and may be automatically programmed for sequential changes in speed, including control of the acceleration and deceleration (17).

Preparation ultracentrifuges are guaranteed for several billion revolutions and can be rebuilt using relatively few parts. Among the great number of rotors available are batch rotors and those accepting feed and discharging centrate continuously during rotation. Batch rotors include angle and swinging-bucket types as well as those having vertical tubes parallel to the axis of rotation, which present a very short sedimenting distance and time requirement. Swinging-bucket rotors are also used for density-gradient separations or volume evaluation of the settled cake.

Separation by selective sedimentation on the basis of size and density of the particles may be satisfactory for polydisperse particle systems. However, the cake contains a range of material depending on its starting position in the container. Selectivity of separation can be improved by introducing the sample near the surface after the container is up to speed. Reslurrying and recentrifuging may be necessary to achieve purer fractions. Isopycnic separation improves initial separation efficiency where particles differ in density. If the density of the medium is intermediate to the range of densities of particles, higher density particles settle, whereas others remain suspended or rise regardless of size.

Zonal Centrifuge. The use of density gradients in centrifuge rotors greatly increases the sharpness of separations and the quantities of material that can be handled. In principle, the density gradient is established normal to the axis of rotation of the rotor and the highest density is located at the outer radius of the rotor. Low molecular weight solutes such as cesium chloride, sucrose, or potassium citrate, which are compatible with many systems in solution, are frequently used. A natural gradient may be formed by introducing a homogeneous solution and centrifuging for long periods of time. Continuous or step gradients may also be formed by introducing successive layers of solution, the composition of which varies continuously or stepwise from low to high density, where the latter displaces the former toward the center of the rotor (17–28) (Fig. 10).

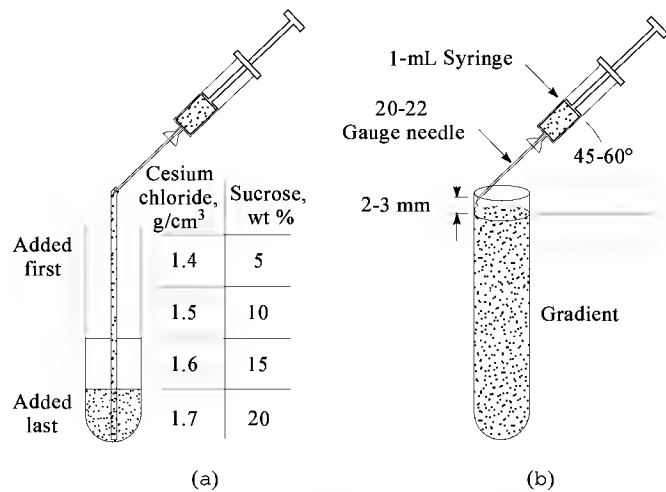


Fig. 10. (a) Forming a gradient and (b) applying the sample to the gradient before inserting tubes in a centrifuge rotor.

In the simpler rotors using batch containers having swinging, angle, or vertical tubes, the gradient is introduced while the rotor is at rest and then accelerated to speed. The gradient shows relatively little mixing. Slowing the rotor gradually at the end of the run allows retention of the gradient and permits collection of the material banded isopycnicly (Fig. 11).

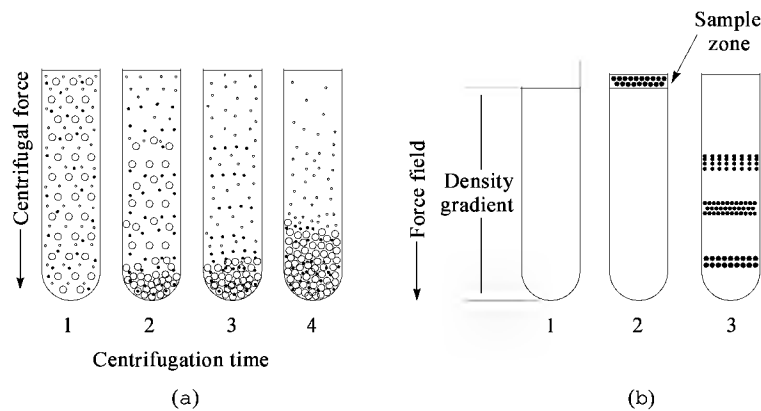


Fig. 11. (a) Differential centrifugation (pelleting), where time 1 < time 2 < time 3 < time. (b) Rate zonal separation in a swinging-bucket rotor, where tube 1 represents the density gradient solution, tube 2 the sample plus the gradient, and tube 3 the separation of sample particles under a centrifugal force, where the particles move at differing rates, depending on their mass.

More sophisticated rotors can be loaded with gradient and sample while rotating. When the batch is finished or the bands are sufficiently loaded with material, the bowl may be stopped slowly and the reoriented layers displaced under static conditions. Rotors may also be designed to establish gradients and isopycnic bands of sample and then be unloaded dynamically by introducing a dense solution near the edge of the rotor as shown in Figure 12.

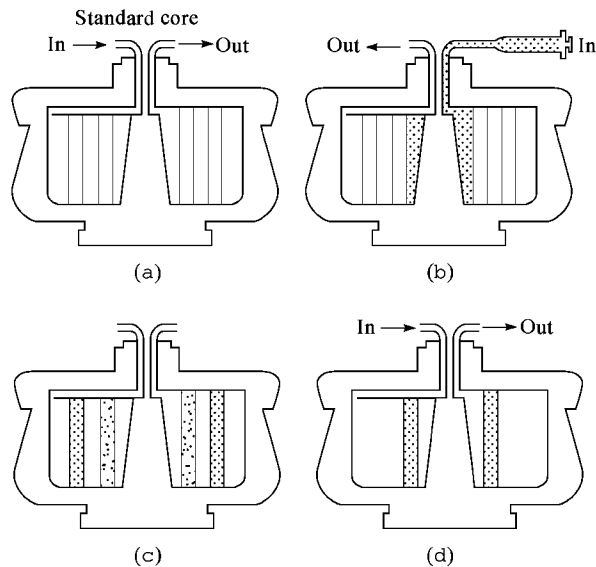


Fig. 12. Dynamic loading and unloading of a zonal rotor. (a) Gradient is loaded while rotor is spinning at 2000 rpm; (b) a sample is injected at 2000 rpm, followed by injection of overlay; (c) particles separated when the rotor is running at speed; and (d) contents are unloaded by introducing a dense solution at the rotor edge, displacing fractions at the center.

Particles in the gradient may be separated on the basis of sedimentation rate; a sample introduced at the top of the preformed gradient settles according to density and size of particles, but the run is terminated before the heaviest particles reach the bottom of the tube. If the density of all the particles lies within the range of the density limits of the gradient, and the run is not terminated until all particles have reached an equilibrium position in the density field, equilibrium separation takes place. The steepness of the gradient can be varied to match the breadth of particle densities in the sample.

Rotors are made of titanium or aluminum and may be cylindrical or bowl-shaped (see Fig. 12). Larger bowls reach 100,000 G; smaller units reach 250,000 G. The tubular rotors permit feed rates up to 60 L h at 150,000 G or 120 L h in a larger unit at 90,000 G. Such centrifuges may be used to separate relatively large quantities of viral material from larger quantities of cellular and subcellular matter, as, for example, in the production of vaccines (see VACCINE TECHNOLOGY).

Tubular Centrifuges. Tubular centrifuges (Fig. 13) separate liquid–liquid mixtures or clarify liquid–solid mixtures having less than 1% solids content and fine particles. Liquid is discharged continuously, whereas solids are removed manually when sufficient bowl cake has accumulated. For industrial use, the cylindrical bowls are 100–180 mm in diameter with length-to-diameter ratios ranging from 4–8. Bowl speeds up to 17,000 rpm generate centrifugal accelerations up to 20,000 G at the bowl wall (14). Because of the small bowl diameter, however, Σ_T values according to equation 11 result in flow rates in the range of 0.1–4 m³ h (0.5–16 gpm). The tubular centrifuge handles low to medium flows and theoretical particle settling velocities in the range of 5×10^{-6} to 5×10^{-5} cm s. Designs for up to 101.3 kPa (1 atm) and 300°C are available. Clean in place (CIP) and sterilize in place (SIP) are available. Drive motors of 1.5–7.5 kW are used.

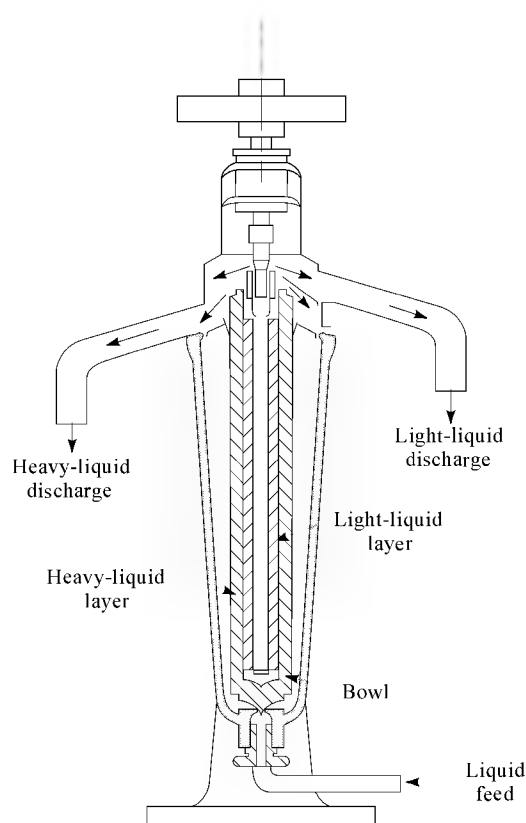


Fig. 13. A tubular centrifuge.

Courtesy of Alfa Laval.

The laboratory tubular centrifuge is similar to the industrial model. It operates with a motor or turbine drive at speeds to 50,000 rpm, generating 65,000 G at the latter speed in the 4.5 cm diameter bowl. The nominal capacity range is 30–2400 cm^3 min. This centrifuge is uniquely capable of separating far finer particles than any other production centrifugation equipment except the bottle centrifuge. It is widely used in the production of flu virus.

A long, hollow, cylindrical bowl is suspended by a flexible spindle and driven from the top as shown in Figure 13. Axial ribs in the bowl ensure full acceleration of the liquid during its short time in the bowl. Feed is jetted into the bottom of the bowl and clarified liquid overflows at the top, leaving deposited solids as compacted cake on the bowl wall. The clarifying performance of the bowl is reduced as the deposited cake decreases the effective outer radius of the bowl in accordance with equation 11. Consequently, cake capacity of the industrial model is limited to 0.1–10 L. For liquid–liquid separation, the interface position (eq. 26) is determined by selection of ring-dam diameter or by the length of a hollow nozzle-type screw dam.

The tubular centrifuge was long used for the purification of contaminated lubricating oils because of the high centrifugal force developed and the simplicity of its operation. Colloidal carbon and moisture are removed from transformer oils to maintain dielectric strength; carbon and acid sludges are removed from diesel engine lubricating oils; and water and solid contaminants are removed from steam turbine lubricating oil. Polishing operations include the removal of small quantities of solids in the clarification of varnish, cider, fruit juices (qv), and even highly viscous chicle. In vegetable oil refining, oil losses in the semisolid soap stock are kept low by compaction of the soap phase under high centrifugal force. Automatic disk centrifuges which do not require manual solids unloading have largely replaced the batch-operating tubular. The laboratory tubular centrifuge is used to recover fine solids in batch preparations too large for bottle centrifuge separation, to estimate scale-up rates in larger centrifuges, and to analyze particle size distributions involving settling rates too low for feasible gravity sedimentation (19). Modern machines having variable-frequency drives are available (14).

Disk Centrifuge. Centrifuges that channel feed through a large number of conical disks to facilitate separation combine high flow rates with high theoretical capacity factors (see Fig. 8). For industrial units flow rates up to 250 m^3 h (1100 gpm) can be obtained on easy separations, and theoretical settling velocities may range from 8×10^{-6} to about 5×10^{-5} cm s. Both liquid–liquid and liquid–solid separations are performed using feed solids concentration below 15% and small particle sizes. As seen from equation 15, the theoretical capacity factor depends on the number of disks, which is limited by the height of the disk stack. The performance is proportional to the cube of the disk diameter.

Several of the assumptions in the development of Σ_D do not apply in practice (20), and mathematical representation of the actual flow pattern has been difficult to achieve. Computer studies of the flow to and between the disks, as well as experimental analysis of flow patterns within the disk stock, have improved the effectiveness of disks. Not all disk machine designs can be compared using Σ_D alone. Details such as type and size disk spacer, number of spacers, and location of the feed holes with respect to the spacers and solids discharging ports are very important. Also the method of feed acceleration is especially important in liquid–liquid separations and in the presence of fragile solids. If the disk centrifuge design is geometrically similar, scaling up by Σ_D from one speed and size of stack to another is reasonably accurate.

The outstanding feature of the disk bowl design is a stack of thin cones, commonly referred to as disks, which are separated by thin spacers. These are so arranged that the mixture to be clarified must pass through the disk stack before discharge. The resulting stratification of the liquid medium greatly reduces the sedimenting distance required before a particle reaches a solid surface and can be considered removed from the process stream. The angle of the cones to the axis of rotation is great enough to ensure that solid particles deposited on the surfaces slide, either individually, or as a concentrated phase according to the difference between their density and that of the medium.

The general flow patterns for a liquid–liquid separator and a recycle clarifier, respectively, are illustrated in Figure 14. Feed enters near the center of the bowl from either the top or the bottom, depending on the support, and is accelerated by vanes to the radius at which it enters the disk stack. When the disk stack is used for one phase, as in clarification or classification, the feed is distributed to the stock through the zone between the outer edges of the disks and the bowl wall. The clarified medium is discharged at a relatively small radius, generally at the top of the bowl. For a liquid–liquid separation, with or without solids, feed is distributed by a number of feed channels. The interface of the two coalesced liquid phases is located at the disk feed holes, by appropriate selection (eq. 26) of the heavy- and light-phase discharge radii. During handling of the two liquids, the heavier moves toward the edge of the disks and the lighter moves inward. Separate channels in the bowl and separate cover compartments segregate the discharges.

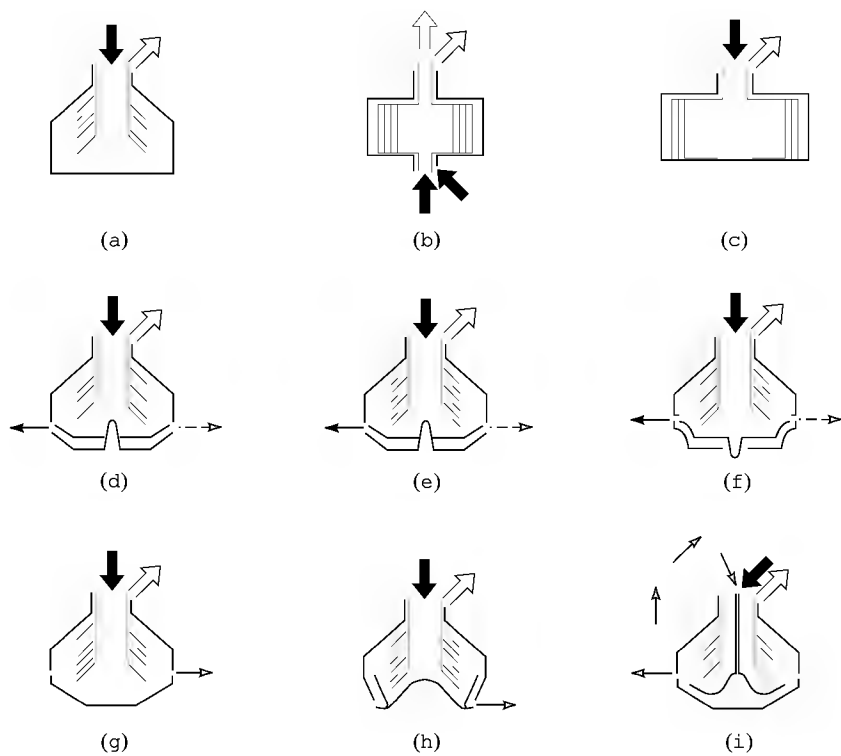


Fig. 14. Disk centrifuge bowls, where bowl diameters range from 10–90 cm; feed flow rates from 0.06×10^{-3} to $38 \times 10^{-3} \text{ m}^3 \text{ s}^{-1}$ (1–600 gpm); and operating temperatures from 10–90°C, and () represents the main inlet, () the main outlet, () continuous flow solids, () intermittent flow solids, and () auxiliary liquid. (a–c) Solid wall bowl: base design, spiral cylinder inserts, and cylinder inserts, respectively; (d–f) bowl for intermittent solids discharges: base design, radial peripheral parts, and shoe; peripheral parts and shoe with nozzles for both continual and intermittent solids discharge; and axial peripheral parts and shoe, respectively; and (g–i) bowl with nozzles: base design and peripheral nozzles, nozzles at reduced diameter, and peripheral nozzles and solids circulation, respectively.

Solids in either phase are sedimented to the underside of the disks and slide outward along the surfaces because of their density. The aggregated solids must move by free settling from the outer edges of the disks to the bowl wall; some may be reentrained into new feed material, and carried into the disk stack, which accounts in part for actual performance falling short of theoretical prediction.

Commonly used with disk centrifuges are centripetal pumps (Fig. 15) that discharge the clarified liquid phases under pressures up to 0.7 MPa (100 psi) at reduced aeration, and scoop the rotating liquid out by using a stationary impeller. Interface location can be altered by varying the centripetal pump-back pressure, allowing interface control without shutdown to change a discharge weir. Centripetal pumps are capable of discharging at rates to 250 $\text{m}^3 \text{ h}^{-1}$ (1100 gpm), and often eliminate the need for a tank and conventional pump (see PUMPS).

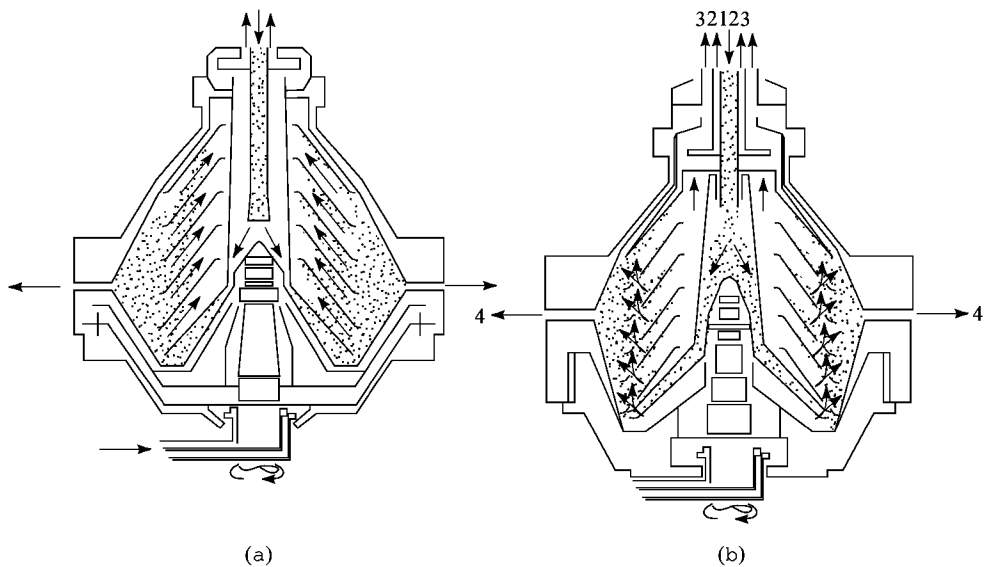


Fig. 15. (a) A clarifier and (b) a purifier of the paring disk-type design, where intermittent discharges of solids are designated by , and 1 represents feed; 2, light phase; 3, heavy phase; and 4, solids.

To maximize cake capacity, the simplest disk centrifuge bowl (Fig. 15a) is designed having a nonperforate bowl wall parallel to the axis. Feed solids should not exceed 0.5%. Bowl diameters of industrial units range from 180–600 mm with operating speeds from 8000 to 4500 rpm; the disks, between 30 and >200, are stacked at spacings of 0.4–6 mm; the half-angle is frequently 35–40° because solids handling is not critical. This type of centrifuge was originally employed for the separation of cream from milk and is still used widely in this field. Other uses include purification of fuel and lubricating oils having a low percentage of solids, separation of wash water from fats and vegetable or fish oils, and removal of moisture and solids from jet fuel. Solids which move readily in plastic flow can be continuously discharged as the heavy phase, for example, in the separation of soap stock from oil in vegetable oil refining.

Continuous discharge of solids, as a slurry, is achieved by sloping the inner walls of the bowl toward a peripheral zone containing between 8–24 orifices, commonly called nozzles, as shown in Figure 14g–i. The nozzles must be spaced closely enough so that the natural angle of repose of the solids deposited between the nozzles does not cause a buildup of cake to reach into the disk stack and interfere with clarification. The size of the nozzles is limited because the fluid pressure at the wall, which can be 6.9–13.8 MPa (1000–2000 psi), produces high nozzle velocities (see eq. 30). On the other hand,

the nozzles must be large enough to prevent obstruction by individual particles; nozzle diameters at least four times the size of the largest particle are satisfactory. Replaceable, wear-resistant bushings in the nozzles provide orifices in the range of 0.8–2.5 mm. From 5–50% of the feed may be discharged with the solids through the nozzles. The upper limit of solids concentration depends on the particle packing characteristics but seldom exceeds 20 times that in the feed. The nozzle flow is directed backward with respect to rotation to recover much of the pressure energy and to reduce centrifuge power (see eq. 29). Bowl diameters range from 100 mm in laboratory units to 900 mm in industrial units and speeds from 12,000–3000 rpm, respectively; power requirements are between 10–150 kW. Centrifugal accelerations to 12,000 *G* are obtained at the wall in the smaller bowls. Feed rates range from 1–136 m³ h (4–600 gpm).

Applications include kaolin clay dewatering, separation of fish oils from press liquor, starch and gluten concentration, clarification of wet-process phosphoric acid, tar sands, and concentrations of yeast, bacteria, and fungi from growth media in protein synthesis (14).

A variant of the continuous-discharge disk centrifuge provides for introduction of a recycle stream. Restrictions on number and size of nozzles sometimes prevent adequate concentration of the discharged solids to satisfy further process requirements. To increase this concentration, a portion of the discharged slurry is returned to the feed, thereby increasing the overall loading of solids in the bowl. A more efficient method, shown in Figure 14i, is to return some discharged slurry through a recycle system to the region of the nozzles where the higher density recycle stream preferentially joins the nozzle flow.

A further modification of the disk centrifuge provides peripheral ports that are opened only intermittently to discharge sludge (see Fig. 14d–f, and Fig. 15). This type is employed for medium quantities of solids (1–4%) for which neither continuous discharge nor batch operation is suitable, and for solids that break down under the shear forces of nozzle discharge and are therefore not suitable for nozzle recycle. Intermittent discharge provides longer holding time and better concentration of solids, often at the expense of decreased disk size and reduced bowl throughput. The frequency and duration of opening can be controlled to discharge very high concentrations of sludge by partial emptying of the bowl.

These centrifuges are available in 180-mm bowl laboratory and 460–600-mm bowl industrial units. Disks are mostly spaced at 0.6–1.0 mm and the half-angle is 40–45°. In the industrial units, bowls with 60–200 or more disks operate at maximum speeds of 4400–6500 rpm and require 10–50 kW. Feed flow rates range from 2.5–114 m³ h (10–500 gpm); temperatures from 30–90°C. Theoretical capacity factors are generally lower than in continuous discharge bowls of the same size. The disk outside diameter is smaller in order to provide solids-holding space of 4–20 L. Applications are limited to free-flowing solids that do not pack, and include recovery of wool grease from wool scouring liquor, orange juice clarification, recovery of soya protein, clarification of animal fats and food extracts, and purification of marine and jet engine fuels and lube oils.

Other modifications have special but more limited applications. A centrifugal bowl may contain, instead of disks, several annular baffles that take the liquid through a labyrinth path before discharge. The multiple cylinders increase cake capacity to as much as 70 L for easily sedimented solids. This centrifuge is used for clarification of food syrups and antibiotics (qv), and for recovery of heavy metallic salts and catalysts (see Fig. 14c).

Continuous Conveyor Discharge Centrifuges. Imperforate bowl conveyor-discharge centrifuges collect solids by sedimentation and continuously discharge both liquid and solid material. These centrifuges have bowl diameters of 150–1400 mm and are essentially tubular shells with a length-to-diameter ratio of 1.5–5.2, as shown in Figure 16. Deposited solids are moved by a helical screw conveyor operating at a differential speed of 0.5–100 rpm with respect to the bowl. Centrifugal fields are lower than in disk or tubular centrifuges because of the conveyor and its associated mechanism. Maximum speeds range from 300–9000 rpm. Figure 8 shows that particles of intermediate settling velocities, such as 1.5 to about 15×10^{-4} cm s, are handled at medium to large flow rates. For clarification, this type of centrifuge recovers medium and coarse particles from feeds at high or low solids concentration. Particle sizes less than about 2 µm are normally not collected without the addition of flocculating agents. For classification of solids, the flow rates are higher than for clarification and the overflow usually contains most of the finer solids. Feed flow rates range from 1 to 136 m³ h (4–600 gpm).

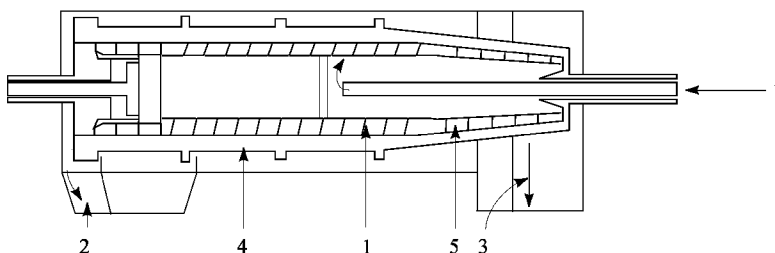


Fig. 16. An imperforate bowl conveyor-discharge centrifuge, where 1 corresponds to feed suspension; 2, to liquid phase; 3, to solid phase; 4, to liquid pool; and 5, to dry beach.

Discharged solids are not as dry as those obtained by centrifugal filtration. Coarse crystals may discharge at 2–10% moisture, ground limestone at 15–20%, and kaolin clay in the filler range (1–10 µm) at 30–35%. Compressible, amorphous, and fibrous materials, such as sewage sludges, can be dewatered to 60–75% moisture, and meat rendering solids at 60–70% moisture plus 6–8% liquid fat. Operating pressures up to 1.03 MPa (150 psi) and temperatures up to 200°C are standard and 300°C is available.

Feed is introduced through a stationary axial tube. Solids adhering as sediment to the bowl wall are removed from the liquid by a conveyor, moved up a sloping beach, and usually discharged at a radius smaller than that of the liquid discharge. Fine and flocculent solids compact under the liquid and show relatively little drainage on the beach. Coarse crystals and fibers do drain on the beach to a low residual moisture. The liquid level in the bowl is maintained by ports adjustable to the desired overflow radius. Considerable variation in design is available for the bowl shell, flight angle and pitch, beach angle and length, conveyor speed, feed position and type, and patterns of liquid and solids movement through the bowl. Bowl shapes having a high l/d exhibit high clarification capacity but may have wetter solids owing to the greater amount of fine particles in the solids discharge.

Bowl designs include countercurrent or concurrent movement of the phases. In countercurrent flow, feed enters near the conical–cylindrical intersection; liquid flows toward the far end of the bowl to discharge over dams; and deposited solids move toward the cake end with the conveyor. In concurrent flow, feed enters at the end away from the cake discharge, liquid and settled solids move in the same axial direction toward the cake end, axial conduits or a skimming device remove centrate from the pond surface, and solids move up the beach. The actual liquid flow between the helical flights is more tangential than axial so the concurrent–countercurrent description is not particularly significant. More important is the problem of introducing the feed stream exactly where the deposited solids are deepest, and the resulting reslurrying in countercurrent designs. Also there is the problem of discharging clarified centrate just where the solids are deepest, and the resulting reslurrying in concurrent designs. Care of feed introduction in countercurrent, and centrate discharge in concurrent, are important factors which determine separation efficiency. Plugging of return tubes and maintenance of the feed–centrate seal are problems unique to concurrent designs. In either type, solutions of flocculating agents may be introduced in the piping before entering the centrifuge, either in the feed tube, the feed acceleration zone, or in the pond after feed acceleration. Owing to the complexity of the chemistry of polyelectrolyte flocculation rate and efficiency, and floc damage owing to turbulence, the best solution for optimum polyelectrolyte use must be determined experimentally. A wash can be applied to the solids on the beach, but efficient rinsing depends on careful design to direct the rinse liquid to the solids. The wash is not collected separately but is discharged with the mother liquor.

A vertical decanter design using a bowl elastically suspended from a spindle with no bottom bearing is often used for high pressure and high temperature applications instead of horizontal axis decanters. Vertical units are easier to seal for pressure operation, and are well suited to accommodate bowl expansion at high temperatures. A vertical, elastically supported rotor as compared to a horizontal design does not offer separation advantages. There

are mechanical advantages, however: the vertical design uses only one seal to three in the horizontal design. Moreover, the process connections are rigid (vertical) rather than flexible (horizontal), thermal expansion is not critical, and noise is lower.

Although Σ_T (eq. 11) indicates a reduced sedimentation performance level at increased liquid depth, this occurs only for coarse solids. The optimum pond depth varies with feed zone design, tendency of deposited solids to redisperse, conveyor differential speed, and particularly the depth of the cake layer required to produce a given solids concentration. Deep ponds are generally more effective for soft, slimy solids because conveying problems may reduce performance level and prevent complete clarification, even at low flow rates. The difficulty with which a scroll centrifuge is able to discharge soft, slippery solids has limited use in the past. To move the solids up the beach the solids must remain at bowl speed, with the helical flights moving the solids inward, against the high G . Often the soft solids slide back into the pond, building up in the bowl and ruining sedimentation performance. Setting the pond level inward of the solids discharge radius and separating the feed portion of the pond from the cone end using a deeply immersed baffle permits the centrifugal pressure head developed to assist the conveyor in discharging the solids. This method is used to thicken secondary sewage sludge from feed, ie, solids concentrations of 0.5–1% to solids concentrations of 5–10%. This thickening reduces total flow to the next stage of treatment by an order of magnitude. Flow rates up to 100 m³ h (440 gpm) have been achieved without the use of polyelectrolytes. A small amount of polyelectrolytes, however, can double this rate.

Dewatering sewage sludge to high solids levels has been achieved by the use of high (5–25 kg t) polyelectrolyte dosage, increased (2000–4000 G) G level, and longer solids residence time in the bowl. Longer residence has been achieved by mechanically restricting the solids near or at the beach and reducing the differential speed to 0.5–4 rpm. The torque between the bowl and conveyor is a good measure of solids dryness, and use of this factor has permitted automatic differential speed control to maintain a constant solid dryness at varying feed concentration and rate. Differential speed control can be obtained by driving the conveyor using a hydraulic motor mounted on the bowl and using an external hydraulic power supply to deliver variable rates of high pressure oil to the rotor by means of a rotary union. Control is also achieved by use of a planetary gearbox mounted on the bowl, the gear box receiving controlled torque, and speed from a variable speed motor or electrical brake. All methods are programmable using microprocessors. Higher torque differential speed controllers have permitted centrifuge capacity and solids dryness to achieve levels of 100 m³ h (440 gpm) at 35–40 wt % solids.

Final dewatering of sewage, especially anaerobically digested sludge, results from the solids almost filling the interior space between the conveyor hub and bowl shell interior. The use of polymers as flocculating agents results in easy separation in spite of a greatly reduced residence time for the centrate. The polymer also conditions the solids so that the pressure resulting from the deep layer of solids and the force needed to move the solids axially and inwardly toward discharge compresses and squeezes additional water from the space between and within the flocs. A drier discharged cake results. Compatibility studies (29) can be made to characterize particular solids. Improvements in dryness is proportional to $G^{1/5}t$, where G is the level at the mean solids radius and t is the solids residence time within the centrifuge. On a scroll decanter having an $l/d = 4.2$, the solids have been found to occupy 75–80% of the bowl. The centrate clarification occurs just before discharging.

Theoretical studies (30) comparing the ability to dewater compressible solids by sedimenting and filtering centrifuges to pressure filters, have shown that at high G levels, scroll decanters produce drier cakes than pressure filtration.

Scroll centrifuges, which include perforated screens, are used to dewater coarser solids to a very high degree. Feed slurry is introduced conventionally into an imperforate scroll machine. The dewatered solids are deposited onto a cylindrical screen with an inside diameter equal to the solids discharge diameter. The conveyor moves the solids axially over the bar screen area where further dewatering and washing takes place. The centrate from the screen is recycled to the feed to further recover fines while the solids pass out of the solids discharge. Typical screen decanter applications are fine coal dewatering and production of purified parazylyene (12).

The capacity of decanters can be limited by any one of several factors (31): centrate clarity, usually a function of Σ ; solids dryness, imposed by requirements of the next process step; conveyor torque, limited by rating of gear box, hydraulic motor, or back-drive capacity; chatter, by torsional instability resulting from stick-slip action of conveyor flights on fusible solids; swallowing capacity, by the ability of conveyor to accept feed without rejection; power, by the rating of drive motor; solids purity, where purity is obtained by rinsing; erosion rate, where abrasive particles are present; size purity, for particle classifications; solids volume, where solids feed concentration is high; fused solids, rate at which pasties are formed; and control, the maximum rate at which process control can be maintained.

Special designs have been offered which include one or more of the following features in vertical or horizontal construction: vapor-tight and pressure-tight enclosures; clean in place (CIP) and sterilize in place (SIP) rotors; three-phase separations having two immiscible liquids and solids; sanitary construction; a wide range of helices, pitches and leads, beach angles, compound beach angles, l/d ratios, pond depth/ d ratios, and ribbed or grooved beaches or bowls; centrate centripetal pump devices; and abrasion protection systems, including complex flight tip geometries of ceramic, carbide, and conventional hard-surfacing.

In clarifying operations, the conveyor discharge centrifuge recovers many types of crystals, meal from fish press liquor, and polymers, such as poly(vinyl chloride) and polyolefins. It is also used to dewater coal (24) and to concentrate solids from flue gas desulfurization sludges. Vertical designs, vapor-tight or under pressure, are applied to terephthalic acid, polypropylene, and catalyst recovery. Classification includes separation of particles over 2 μm from kaolin coating clay, and of particles over about 5 μm in the mill discharge of ground TiO₂; selective recovery of calcium carbonate from lime-treated waste sludges permits calcining and recycling of the lime without an overwhelming recycle load of inert material. The conveyor centrifuge is frequently used to rough out medium and coarse solids before a second separation stage such as a disk centrifuge handling refinery sludges.

Centrifugal Filtration Equipment. The important parameters of centrifugal filtration equipment (4,32) are screen area, level of centrifugal acceleration in the final drainage zone, and cake thickness. The latter affects both residence time and volumetric throughput rate. As indicated by equation 19, the particle size of the solids and the kinematic viscosity of the mother liquor also strongly affect the final moisture content. A limited correlation has been developed for the performance of perforate basket and conveyor discharge conical screen bowls, but the range of materials handled and the complexity of the drainage and washing operations do not lend themselves to broad correlations. The variables of correlation may be useful in a particular study, especially if more than one type of centrifuge is involved. An example of centrifugal filtration is the recovery of salt crystals from a mother liquor of about 3×10^{-6} m² s (3 cSt) viscosity; the crystal size is 170 μm at the 15% level and drained cake bulk density about 1.5×10^3 kg m⁻³ (95 lb ft⁻³). The cone screen is 25.4 cm at the larger diameter and the automatic cycle basket centrifuge is 68.6 cm in diameter; other parameters for final values of $q = 7.6\%$ in both centrifuges are given in Table 1.

Table 1. Centrifuge Parameter Values for $q = 7.6\%$

Properties	Centrifuge	
	Cone screen	Automatic basket
G	1440	865
bowl		
diameter, cm	25.4	68.6
speed, rpm	3180	1500
$d(G)^{1/2} \nu$	3.750×10	2.880×10
time, s		
feed and spin		12
spin after rinse		27
drain time	0.6^a	39
rinse ^b		10
unload and screen rinse		4

cycle time		53
$qt^{1/2}, \text{ }^{\circ}\text{0}\cdot\text{s}^{1/2}$	5.89	47.46
$b, \text{ cm}$	0.51	3.81
screen area, cm^2	930	7675
approximate cake rate, kg/s	1.19	0.84

^a Differential speed 60 rpm, 5–8 turn per helical flight.

^b Rinse is shown only in the basket centrifuge because rinse efficiency in the cone screen is fairly low compared to the basket. The latter may be selected for the application if good rinsing is needed, and the conveyor-discharge conical screen centrifuge selected if no rinsing is necessary.

A conveniently expressed coordinate for plotting filtration performance is the drainage number, $\bar{d}(G)^{1/2}/\nu$, where $\bar{d}$ is the mean particle diameter in micrometers, ν is the kinematic viscosity of the mother liquor in m^2/s (Stokes $\times 10^{-4}$) at the drainage temperature, and $G = \omega^2 r/g$; r is the largest screen radius in a conical bowl. Because the final moisture content of a cake is closely related to the finest 10–15% fraction of the solids and is almost independent of the coarser material, it is suggested that $\bar{d}$ be used at the 15% cumulative weight level of the particle size distribution instead of the usual 50% point.

The other coordinate is $qt^{1/2}$, where q is the percentage of final moisture on the discharged cake, as the volume of mother liquor per unit volume of solids, and t is the drainage time in seconds. A weight ratio may be used for q , but the volume ratio makes the function more universal by eliminating densities. For a conveyor discharge conical screen bowl, the helical conveyor is assumed to control the residence time of the solids, so that time becomes the number of turns of helical flight around the conveyor hub divided by the differential speed between the conveyor and the bowl. For a pusher centrifuge, t is the retention time on the screen as controlled by length and diameter of screen, thickness of cake, and frequency and length of stroke of the cake. In a basket centrifuge, dead and unload times should not be included in the calculation of drain time. Bulk drainage is completed so quickly that film drainage is usually controlling. Thus, drain time t is approximated by the sum of feeding time, spin time prior to rinsing, and spin time after rinsing but prior to unloading according to the theory of cyclical centrifuges (33). Filtration correlations generally show a spread as a function of cake thickness. Conical screen bowls characteristically have short residence times and achieve good drainage by maintaining thin layers of cake. Smaller perforate baskets operate having cakes 5–10 cm in thickness, whereas larger baskets may carry cakes up to 15-cm thick. Pusher centrifuges and high speed peeler baskets may handle cakes ranging in thickness from 5–20 cm.

Perforate Basket Centrifuges. The simplest and most common form of centrifugal filter is a perforate-wall basket centrifuge, consisting of a cylindrical bowl having a diameter ranging from about 100–2400 mm and a diameter-to-height ratio ranging from 1–3. The wall is perforated with a large number of holes, more than adequate for the drainage of most liquid loads, and is lined with a filter medium. In the simplest case, the medium is a single layer of fabric or metal cloth or screen. In high speed basket centrifuges, one or more backup screens of relatively large mesh support a finer mesh filter surface. The method of discharging accumulated solids distinguishes three types of basket centrifuge: those that are stopped for discharge, those that are decelerated to a very low speed for discharge, and those that discharge at full speed (34).

Basket centrifuges that must be stopped for discharge are available in many sizes. The bowls are usually supported on a vertical spindle. Designs vary from a 300-mm diameter basket of 30-L cake capacity (0.4-kW motor, 2100 rpm, and about 800 G) to a 1500-mm diameter basket having 0.5- m^3 cake capacity (14-kW motor, 600 rpm, and 300 G). Basket cake volumes are always nominal and must be modified according to cake density. Construction materials include carbon or stainless steel, Monel, Inconel, titanium, and a variety of rubber and plastic coatings. Normal operation includes pressures up to 35 kPa (5 psi) and temperatures up to 180°C. This type of centrifuge is used if a variety of materials must be filtered in small batches, if equipment must be sterilized between batches, or if the production rate is too low to warrant more automation. These centrifuges are also used in removing liquid from crystalline materials, in drying bulk materials such as raw leafy vegetables, and in clarifying process liquids and waste streams. Cycle times are above 10 min. Large (dia 1300–1500 mm) baskets operating at 700–800 rpm may use an inner perforated container which mounts in the bowl but is removable for bulk loading outside the centrifuge. Such units are well suited to laundry and dyeing purposes and for dewatering of textiles, yarn, raw stock, feathers, and hair. Particle sizes range from very fine to 500 μm . Feed flow rates vary from 1–20 m^3/h (4.5–90 gpm). A syphon to control the rate of filtrate removal improves dewatering (35).

Improved control systems and rising labor costs have led to manually or automatically controlled cycles with mechanical unloading at reduced speed. Baskets typically load at low to medium speed, accelerate to 900–1800 rpm for drainage and washing, decelerate to 35–75 rpm for mechanical unloading, and then start the cycle again. Cycle times range from 2–6 min. Cycles of 30 min for slow drainage or multiple rinses are common. Both a top-driven suspended bowl and underdriven bowl with three-point casing suspension designs are available. The cake is discharged in 20–120 s by means of a single or multiple plow that leaves a heel of cake on the filter medium. The heel can be completely removed with the help of a plastic-tipped plow and a perforated protecting plate. Filter media vary from perforated plates having 3-mm holes to 37- μm (400-mesh) Dutch twill. Baskets are always bottom discharge; a valve mechanism may be used to seal the bottom if the basket is fed with the whole charge at one time, so that an appreciable liquid layer develops.

The most fully automated basket or peeler centrifuge feeds and discharges at spin speed, and normally operates at cycle times of less than 3 min at pressures up to 1.03 MPa (150 psi) and temperatures from 70 to 120°C, and in special cases to 350°C. It is primarily used for materials draining freely in a high centrifugal field, for medium tonnages, and for multiple rinses where nearly complete segregation of the rinses and mother liquor is desired. Ideally, the feed should have a constant composition and high concentration. For this purpose, a gravity slurry concentrating tank or cyclone is often installed in front of the centrifuge. The bowl rotates on a horizontal axis with a metal screen as the filter medium and discharges solids by cutting them out with a hydraulically operated knife. Bowls are 300–1200 mm in diameter and handle loads of 28–170 L (1–6 ft^3) of cake at bowl speeds of 1000–2500 rpm, producing maximum centrifugal fields of about 1250 G . The solids discharge through a front chute, which simplifies sealing against pressure. A distributor riding on the cake during feeding maintains uniform cake thickness, gives better bowl balance, and improves washing efficiency. Power requirements range from 15–150 kW. Materials of construction include carbon or stainless steel and Monel. The high speed during feeding and discharge may cause deformation of particles and breakage of crystals, whereas the heel of cake left on the screen may lose permeability through glazing or plugging with fractured fines. The heel can be conditioned by suitable washes after one or more cycles. Table 2 lists two process cycles.

Table 2. Basket Centrifuge Operation on Free- and Slow-Draining Particles

Operation	Type of crystal	
	Free-draining ^a	Slow-draining ^b
solids handling rate, t/h	20–24	1.5–2
conditioning rinse, s	1	0
feeding time, s	7	25
wash time, s	5	25
drain time, s	12	30
unloading time, s	2	6
total cycle, s	27	86

^a For example, ammonium sulfate.

^b Requiring several rinses, eg, polyolefins.

This type of centrifuge is also used on borax and boric acid, *p*-xylene, sodium bicarbonate, and sodium chloride from glycerol or electrolytic caustic,

in addition to dewatering of various products, eg, potato starch, or dewatering and washing slurries, eg, calcium hypochlorite.

Inverting Filter Centrifuge. Another batch automatic horizontal perforated bowl centrifuge inverts the flexible filter to discharge the solids. Feed slurry may be deposited on the inside surface of a cloth filter, with the bucket end completely closed. When the interior is full of dewatered solids, the bowl is decelerated to a slow speed and piston and closure plates move axially, inverting the filter cloth so that the solids reside on the outside diameter of the cloth. Very little residual material remains on the filter cloth surface for the next cycle. Centrifuges are available with diameters of 300–1300 mm, and operate at 720–1920 g. Filter cake rates of 75–250 kg h are achieved at excellent rinsing efficiencies and low crystal breakage. Drives are electric or hydraulic, with the axial movement hydraulically actuated via a rotary seal. Power demands are low, from 3 kW for the smallest to 100 kW for the largest.

Continuous Cylindrical Screen Centrifuges. Continuous filtering centrifuges are used for very fast draining that do not require extremely dry final products. Rinsing efficiency varies considerably; power requirements are usually low; initial slurry concentrations can be somewhat more variable and not as high as for the high speed basket. Continuous centrifugal filters are equipped with either a cylindrical or a conical screen. Both types are made without a retaining lip on the solids discharge end of the bowl and employ various methods to move the solids through the bowl.

The cylindrical screen centrifuge deposits solids at one end of the bowl in a layer 6–80 mm thick and pushes the annular ring of cake axially through the bowl by means of a reciprocating piston (36,37). Washes are collected in separate sections of the casing but are not as distinctly separated as with basket centrifuge sequential operation. Drained solids at the end of the bowl are thrown into a casing that is separated by baffles from the liquid discharge zones. Bowls rotate on a horizontal axis, range in diameter from 200–1200 mm, and have capacities of 1–25 t of solids h. Centrifugal fields of 300–600 G are common. To reduce the fines loss and to facilitate movement of the cake on the screen, maximum speeds are rarely used. Power requirements range from 4–60 kW. The cylindrical screens are generally of the bar screen type, with 0.1–0.5-mm spacing. The reciprocating piston operates at 20–100 strokes min, with stroke lengths up to 80 mm. The thickness of the cake (max about 80 mm) depends on the packing and draining characteristics, the bulking tendency of the layer in front of the pusher, and the frictional resistance of the cake on the screen. Friction also depends on the construction of the bar screen.

The feed slurry is introduced by gravity flow of 20–200 t h or screw conveyor to an imperforate distributing cone that deposits the slurry at its original concentration immediately in front of the pusher, as shown in Figure 17. The cake must not buckle at this point, so slurries of 40° concentration are generally necessary. Because fast drainage is required, feed particle size should exceed 150 μm; medium and coarse crystals and granules or fibrous solids can be handled in this type of centrifuge. Crystal breakage is low within the basket but some breakage does occur during discharge.

To handle materials that form a soft or plastic cake or have a high frictional resistance, the cylindrical screen may consist of two to six steps with successively larger diameters, as shown in Figure 17. Alternate steps reciprocate with the piston. Thus, the cake is pushed across only a short length of screen before redistribution on the next step at a slightly larger diameter. Drainage and washing efficiency are increased by redistribution of the cake under these conditions. This type of centrifuge is used on sugar (qv), where the high viscosity of the mother liquor causes slow drainage, a high degree of plasticity in the partially drained cake, and poor penetration of wash liquor.

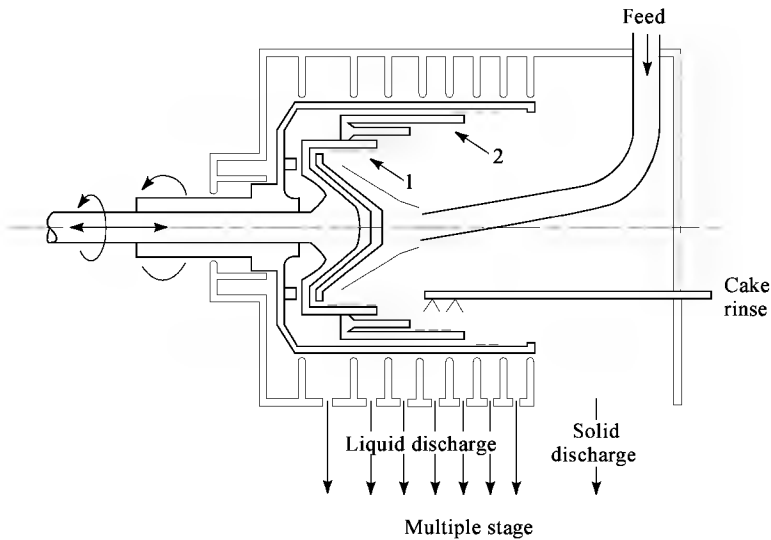


Fig. 17. Multiple-stage pusher centrifuge. Screens 1 and 2 reciprocate with pusher.

Another type of continuous, cylindrical screen centrifuge discharges the cake by moving it axially through the bowl with a helical conveyor. Crystal breakage through conveyor action is greater than with the pusher-type mechanism. Applications include the handling of copper as trisodium phosphate and purification of paraxylene.

Conical Screen Centrifuge. In conical screen centrifuges the angle of the bowl causes or assists the cake to move axially and redistributes it in an increasingly thin layer which improves drainage characteristics. The feed slurry is deposited at the small end of the cone, where most drainage occurs. The drained solids are discharged from the large end, which has no retaining ring. Screens generally have 0.08–1.5-mm slots or perforated plate holes. Screens less than about 0.25-mm thick require a backup screen to extend screen life. There are three types of conical screen centrifuges: those that are self-discharging, those that discharge by means of a helical conveyor (Fig. 18), and those that apply an axial vibration or oscillation to the bowl or the bowl and casing.

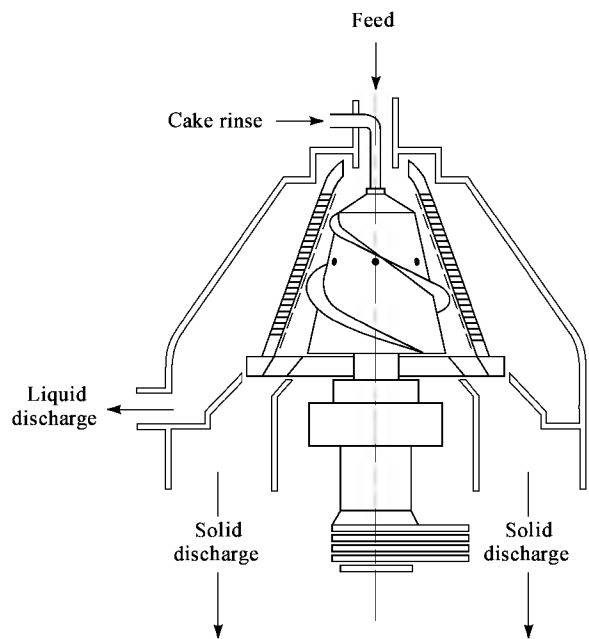


Fig. 18. Conical screen conveyor discharge centrifuge.

In its simplest form, the cone angle is slightly larger than the angles of repose of the solids at any stage in their drying cycle. Some bowls are made with two angles, such that the more shallow is at the small end, and the steeper, where solids concentrate, is higher and at the large end. Horizontal or vertical bowls of 500–1000-mm large-end diameter operate at speeds up to 2600 rpm, producing up to 2500 *G*; cone angles vary from 20–35° and selection of the proper cone angle and suitable screen surface is critical for each application. Temperature control of viscous feeds is necessary to maintain proper distribution and drainage. Feed slurry, usually under gravity flow, is introduced at the small end of the basket, where it is accelerated and spread evenly over the periphery of the screen. Viscous sugar massecuites are successfully handled at capacities of 2–7 t/h on a 750-mm diameter basket. Loss of fines is greater than on the automatic basket centrifuges, but improved rinse performance increases sugar purity. Rinsing efficiency is not generally high on this type of cone and segregation of rinse liquor is incomplete. This centrifuge is also used for the drying of crystalline materials such as ammonium sulfate and separation or dewatering of fine fiber in wet corn milling.

Conical screens with a helical conveyor turning slightly faster than the bowl can handle a variety of materials. The vertical baskets are underdriven and the larger diameter is downward, as shown in Figure 18. Horizontal axis designs are also available. Feed is introduced at the small end and the rate at which solids move through the drainage zone is controlled to some degree by the differential speed of the conveyor. Bowls have large-end diameters of 250–500 mm, and lengths about two-thirds of the large diameter; bowl angles range from 10–20°, and operating speeds are 2500–3800 rpm, giving up to 3500 *G*. Solids capacities range from 1–30 t/h, with feed capacities of 1–15 m³/h (260–4000 gph). Centrifuge casings may be vapor-tight but are not intended for pressure operation; maximum temperature is about 150°C. Power requirements, low for the tonnages handled, range from 7–30 kW. Applications include dewatering of medium and coarse crystals, deoiling of proteinaceous solids, and removal of solids from fruit and vegetable pulps and other food slurries.

The third type of cone screen centrifuge operates with bowl angles of 13–18° and assists solids discharge by a vibratory motion of the bowl or bowl and casing. These units usually have underdriven bowls having 500–1100-mm larger diameter at the upper end; horizontal axis designs are also available where diameter-to-length ratios range from 1–2. Operating speeds are normally 300–500 rpm, and solids capacities range from 25–150 t/h. Pressurized units are not available, and operating temperatures range to 100°C. Power requirements are 15–35 kW. Bar screens are frequently used, and applications are largely on coal dewatering; particle sizes from 30 mm down to 0.25 mm (60 mesh) are easily handled. These centrifuges are also used in dewatering of potash and other crystalline solids.

Gas Centrifugal Separation

Highly developed centrifuges are used to enrich uranium for nuclear application (see NUCLEAR REACTORS; URANIUM AND URANIUM COMPOUNDS). Gaseous uranium hexafluoride, UF₆, is introduced into a very high speed tubular rotor, causing the lighter ²³⁵U-fraction to separate from that of the heavier ²³⁸U. The enrichment that is achieved using one centrifuge is not large enough, so in a commercial plant, the centrifuges are arranged in cascades, where the enriched fraction passes up the cascade for additional enrichment stages, while the depleted fraction passes down the cascade. Each stage of enrichment requires many centrifuges to multiply the total capacity, and many stages to result in an adequately enriched product. The feed has 0.70% ²³⁵U; the product has 3–4% ²³⁵U. This ultimately results in thousands of individual centrifuges, interconnected to each other in the cascade. Extensive successful development has taken place to improve the performance of each centrifuge and reduce the total number of gas centrifuges needed to achieve a satisfactory product at an economic cost per unit of separation power, ie, kg SW/a, where SW = separative work in kg/yr.

A gas centrifuge separative power, *U*, can be expressed as

$$U = \frac{Dp}{RT} \left(\frac{\Delta M \omega^2 r^2}{2 RT} \right)^2 \frac{\pi l}{2} f$$

(37)

where *D*, *p*, and *T* are the gas properties, density, pressure, and temperature, respectively; *R* is the ideal gas constant; Δ*M* is the difference in the molecular weight of the gaseous isotopes; ω²*r*² = surface velocity *v*² of the rotor; *l* = length of the rotor; and *f* = flow factor, which is a function for the actual gas flow pattern within the centrifuge rotor.

In order to increase the separative power of a gas centrifuge, the surface velocity must be as high as possible, the length of the rotor as long as possible, and the gas flow within the rotor introduced and withdrawn from optimum locations, to result in maximum enrichment. Figure 19 is a schematic of a gas centrifuge which operates at surface speeds of 600 m/s, has a diameter of 50 cm (19.7 in.) (147,000 *G*), a length of 500 cm (16.4 ft), and rotates in a vacuum contained within a stationary housing. Owing to the slender shape (*l/d* = 10), the required high speed is well above the first natural frequency of the rotor (even above the third critical). Upper and lower magnetic bearings locate and damp vibration during operation and when passing through the critical speeds. The rotor hub is driven by a rotating magnetic field generated by the stationary stator. The weight is magnetically reduced, with only a needle bearing in the bottom to support and locate the rotor.

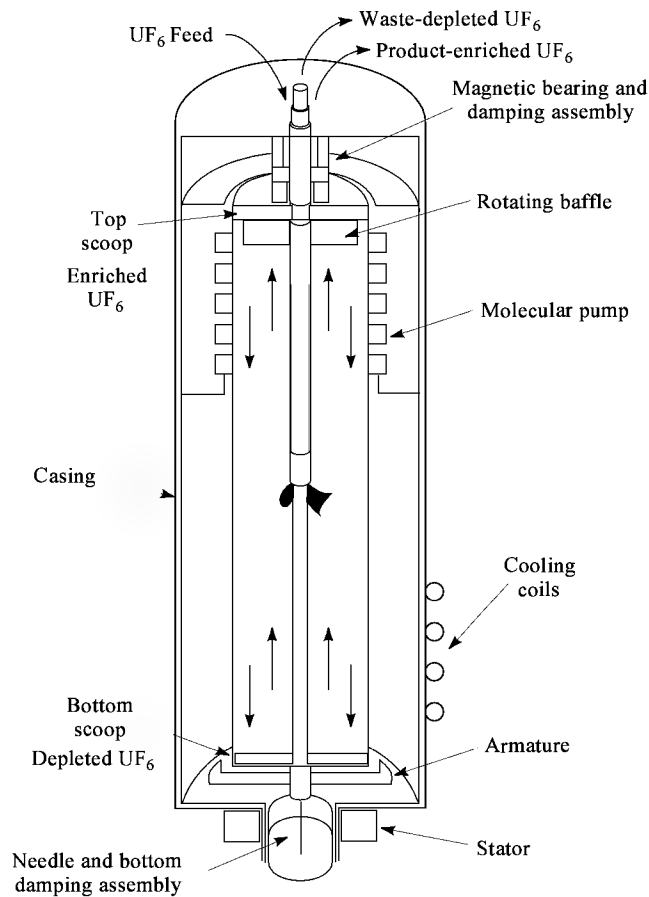


Fig. 19. The modern gas centrifuge and its main components.

The feed gas is introduced near the rotor axis. Enriched and depleted gases are extracted by stationary pitot-like scoops. The location and shape of these tubes, and the baffles within the rotor, greatly effect the gas flow which recirculates within the rotor, reaching enrichment equilibrium at a given feed rate. A vacuum is maintained around most of the rotor. The UF₆ leakage around the stationary axial post is confined to the top of the case by the use of a molecular pump.

Gas centrifuge effectiveness has increased 10–15 times from the 1975 pilot plant owing to increases in speed from stronger bowl (38–41) materials and dynamic bearing and dampening systems. A better understanding of the gas flow within the bowl has also contributed. It is expected that the next generation of gas centrifuges, installed in the same footprint, should have two to three times the separation power by using carbon fiber composites (surface speed > 700 m/s) and further refinement of the feed and extraction devices. Commercial gas centrifuge enrichment plants exist in England, Holland, Germany, and Japan.

NOMENCLATURE	
Symbol	Definition
$\mathcal{A}e_{3/4}$	the cylindrical area at 3/4 of the bowl wall radius
a	distance; space between adjacent disks
C	discharge coefficient
c	concentration
d	diameter of particle; particle size
$\overline{d}$	mean particle diameter
e	efficiency factor
E	energy
E_{rel}	relative performance
G	level at the mean solids radius
G	ratio of centrifugal acceleration to acceleration gravity
g	acceleration of gravity
h	cake thickness; height
K	permeability
k	experimental coefficient in eq. 17
k'	experimental coefficient in eq. 19
l	length, particularly from feed zone to centrate discharge
m	torque
n	number of disks
P	power
p	pressure
q	volume of undrained liquid (mother liquor) unit volume of solid, %
q_{∞}	volume of undrained liquid (mother liquor) unit volume of solid at infinite time, %
Q	flow rate, volumetric
r	radius
s	external surface area weight of solid
s'	surface area volume of cake
S	fraction of void volume occupied by liquid at time, t
S^{∞}	fraction of void volume occupied by liquid at infinite time
t	time during which material is exposed to separation effect
t'	time at which free liquid enters the cake in eq. 18
T	torque

v	velocity
V	volume
α	cake resistance constant in pressure filtration, length / volume
δ	mass density = mass / unit volume
Δ	difference, particularly of density
ϵ	void fraction
μ	viscosity
ω	angular velocity of rotation motion
Ψ	wetting angle
ϕ	angle between direction of the nozzle and tangent to circle intersecting nozzle axis at discharge section
σ	surface tension or material stress
Σ	theoretical capacity factor
θ	half-included angle of disks
ν	kinematic viscosity
<i>Subscripts</i>	
B	bottle centrifuge
BD	back-drive
C	cake, or conveyor torque
D	disk centrifuge
f	film flow
F	friction power
g	settling velocity of particle in gravity field
h	heavy phase
i	interface
l	free surface of liquid or light phase
L	liquid medium
m	mean value
n	nozzle
M	filter medium
N	nozzle
P	process power
p	process
S	solid medium
s	settling velocity of particle in centrifugal field
W	windage power

BIBLIOGRAPHY

"Centrifugal Separation" in *ECT* 1st ed., Vol. 3, pp. 501–521, by M. H. Hebb and F. M. Smith, The Sharples Corp.; in *ECT* 2nd ed., Vol. 4, pp. 710–758, by A. C. Lavanchy and F. W. Keith, Jr., The Sharples Co., and J. W. Beams (Gas Centrifugal Separation), University of Virginia; in *ECT* 3rd ed., Vol. 5, pp. 194–233, by A. C. Lavanchy and F. W. Keith, Jr., Pennwalt Corp.

1. C. M. Ambler, *Chem. Eng. Prog.* **48**, 150 (1952).
2. F. W. Keith, Jr. and R. T. Moll in R. A. Young and P. Cheremisinoff, eds., *Wastewater Physical Treatment Processes*, Ann Arbor Science Publishers, Ann Arbor, Mich., scheduled 1978.
3. J. Murkes, *Brit. Chem. Eng.* **14**(12), 636 (1969).
4. H. P. Grace, *Chem. Eng. Prog.* **49**, 427 (1953).
5. J. A. Storrow, *AIChE J.* **3**, 528 (1957).
6. W. Batel, *Chem. Ing. Tech.* **33**, 541 (1961).
7. E. Nenninger, Jr. and J. A. Storrow, *AIChE J.* **4**, 305 (1958).
8. Technical data, Sharples Research Laboratory, 1961.
9. J. O. Maloney, *Ind. Eng. Chem.* **48**, 482 (1956).
10. Technical data, Alfa Laval Separation, Warminster, Pa., May 1993.
11. T. Theodorsen and A. Regier, *Experiments on Drag of Revolving Disk, Cylinders and Streamline Rods at High Speeds*, Report No. 793 National Advisory Committee for Aeronautics 11944, pp. 367–384.
12. Technical data, Alfa Laval Separation, Warminster, Pa., 1984.
13. *Chem. Eng.*, 353–356 (Aug. 1993).
14. H. Axelsson, *Centrifugal Separations—Principles and Techniques*, Alfa Laval Separation, Tumba, Sweden, presented at the Bioprocess Technology Program, University of Virginia, Charlottesville, Va., Oct. 17–25 1991.
15. C. M. Ambler and F. W. Keith, Jr., in A. Weissberger, ed., *Techniques of Chemistry*, 3rd ed., Vol. XII, Wiley-Interscience, New York, 1978, Chapt. VI.
16. P. A. Vesilind, *Treatment and Disposal of Waste Water Sludges*, Ann Arbor Science Publishers, Ann Arbor, Mich., 1974.
17. O. M. Griffith, *Techniques of Preparative, Zonal and Continuous Flow Ultracentrifugation*, Spinco Division of Beckman Instruments, Inc., Palo Alto, Calif., 1975.
18. G. B. Cline in E. S. Perry and C. F. van Oss, eds., *Progress in Separation and Purification*, Wiley-Interscience, New York, 1971, pp. 299–306.
19. T. Lee and C. W. Weber, *Anal. Chem.* **39**, 620 (1967).
20. C. A. Willus and B. Fitch, *Chem. Eng. Prog.* **69**, 73 (Sept. 1973).
21. D. E. Albertson and E. E. Guidi, Jr., *J. WPCF* **41**, 607 (Apr. 1969).
22. F. A. Records, *Chem. Eng.*, 81 (Jan. 1974).
23. R. J. Woolcock, *Filtr. Sep.* **12**, 174 (Mar. / Apr. 1975).
24. J. J. Halloran, No. TIS-5039, *Annual Meeting, Shurry Transport Association*, Houston, Tex., Aug. 24–25, 1976.
25. R. Day, *Chem. Eng. Prog.* **69**, 67 (Sept. 1973).
26. D. F. Kelsall, *Trans. Inst. Chem. Eng. London* **30**, 87 (1952).
27. D. F. Kelsall, *Chem. Eng. Sci.* **2**, 254 (1953).
28. D. Bradley, *The Hydrocyclone*, Pergammon Press, London, 1965.
29. P. A. Vesilund, B. Zhang, *J. WPCF*, **56**(12), 1231–1237 (Dec. 1984).
30. F. M. Tiller, N. B. Hsyung, *Chem. Eng. Prog.*, 20–28 (Aug. 1993).
31. W. Gosile, *German Chem. Eng.* **3** (1980).

32. R. Day, *Chem. Eng.*, 81 (May 13, 1974).

33. F. A. Records, *Chem. Proc. Eng.* **52**, 47 (Nov. 1971).

34. Technical data, Alfa Laval Separation, Warminster, Pa., Sept. 1991.

35. K. Lilley and G. Huhtsch, *Filtr. Sep.* **12**, 70 (Jan.–Feb. 1975).

36. D. K. Baumann and D. B. Todd, *Chem. Eng. Prog.* **69**, 62 (Sept. 1973).

37. P. M. T. Brown, *Chem. Proc. Eng.* **52**, 65 (Nov. 1971).

38. T. T. Edwards and M. D. Holmes, *Nucl. Eng. U.K.* **28**(6), 174–177 (Nov.–Dec. 1987).

39. I. D. Heriot, *Uranium Enrichment by Gas Centrifuge*, EUR 11486EN, for the Commission of the European Communities, Urenco Ltd., Luxembourg, Belgium, 1988.

40. *Fundamentals of Uranium Enrichment, Pt. 2. Methods: Coaseous Diffusion and Gas Centrifuges*, Nuclear Fuel Cycle Training Course, X ITP—269 P2; DE89 011789 U.S. Air Force Technical Application Center Project T 9427 NP DOE; The International Technology Programs Div., Martin Marietta Energy Systems Inc., Oak Ridge, Tenn., May 1989.

41. K. Cohen, *The Theory of Isotope Separation as Applied to Large Scale Production of 235V*, McGraw-Hill Book Co., Inc., New York, 1951.

Alan Letki
R. T. Moll
Alfa Laval Separation Inc.

Leonard Shapiro
Consultant



MAGNETIC SEPARATION

Principles

The application of magnetic separators relies on the behavior of individual particles under the influence of magnetic forces. When exposed to a magnetic field, all materials are affected in some way. Those that are attracted to the magnetic field are designated ferromagnetic; those that are repelled by it are called diamagnetic. Ferromagnetic materials are attracted along lines of magnetic force from points of lower magnetic field intensity to points of higher magnetic field intensity. The particles frequently retain magnetic properties after the applied magnetic field is removed. Diamagnetic materials are repelled along the lines of magnetic force from points of high magnetic intensity to points of lower field intensity. Only limited commercial use of diamagnetic effects has been made.

Table 1. Magnetic Induction Required to Extract Discrete Minerals

Mineral	Magnetic intensity, T ^a	Mineral	Magnetic intensity, T ^a
alabandite	1.5–1.8	limonite	1.6–2.0
ankerite	>1.3–<1.6	maghemite	0.3–0.5
apatite	1.4–1.8	magnetite	≪0.1
bastnasite	1.5–1.7	martite	0.2–0.6
biotite	1.0–1.8	monazite	<1.4–2.0
braunite	1.4–1.8	muscovite	1.5–2.3
chromite	1.0–1.6	olivine (fayalite)	1.1–1.5
chrysocolla	2.0–2.4	pyrochlore	1.2–1.6
columbite	1.2–1.6	pyrolusite	1.5–1.9
columbite–tantalite	1.2–1.6	pyrrhotite	0.1–0.4
davidite	1.2–1.6	renierite	1.4–1.8
epidote	>1.4–2.0	rhodochrosite	1.5–2.0
euxenite	1.6–2.0	rhodonite	1.5–2.0
ferberite	>0.1–0.4	samarskite	1.6–2.0
franklinite	<0.3–<0.5	siderite	<1.0–1.8
garnet	1.2–1.9	staurolite	1.2–1.9
goethite	1.5–1.8	serpentine	>0.4–>1.8
haematite	>1.3–2.0	tantalite	1.2–1.7
hornblende	>1.6–2.0	titaniferrous magnetite	<0.1–0.3
ilmenite	0.8–1.6	tourmaline	1.6–2.0
ilmenorutile	1.5–1.8	uraninite	1.8–2.4
itabirite	0.8–<1.4	wolframite	1.2–1.6
		xenotime	1.1–1.6

^a To convert T to G, multiply by 10⁴.

For practical applications, ferromagnetic materials are described as strongly magnetic (ferromagnetic), magnetic, weakly magnetic, or nonmagnetic. The limits of these groupings are not well defined and can vary in some mineral species. Values indicating the magnetic intensity required to separate a selected grouping of minerals are shown in Table 1. Only a few minerals are ferromagnetic, eg, magnetite. The dividing line between high and low intensity magnetic fields has been selected by some investigators to be one Tesla (10 kG).

Magnetic separation methods are used either to separate valuable minerals from nonmagnetic waste, or magnetic impurities or other valuable magnetic minerals from bulk nonmagnetic values. Magnetically susceptible mineral particles occur as individually discrete particles; as partially liberated particles consisting of two or more minerals, one of which is magnetically susceptible; or as particles stained or coated by a magnetically susceptible mineral such as goethite staining quartz or kaolin.

Minerals normally considered nonmagnetic may be rendered magnetic by elemental substitution of a small amount of a magnetic element in the crystal lattice. Magnetic properties may also be affected by partial alteration in weathering effects.

Definitions. Terms used in magnet evaluation are distinct from those in other processes and are defined as follows. Magnetic flux density refers to the number of magnetic lines of force passing through a unit area and is measured in Tesla (Gauss). One Tesla equals 10,000 Gauss. Magnetic field intensity, ie, the magnetizing force that induces the lines of force in the specific area under consideration, is also measured in Tesla. In addition, the term magnetic field gradient is used and refers to the rate of change of magnetic flux density from areas of low intensity to points of high intensity. This term is measured in Tesla per meter (Gauss per centimeter). Ferrous particles placed in a magnetic field can serve to focus magnetic lines of force. Thus the particles can become sites of high intensity.

Equipment. From the applications standpoint, magnetic equipment falls into one of four broad categories: tramp iron removal, magnetic particle separation and concentration, product cleaning, or eddy current separation on nonmagnetic metallics.

Tramp iron removal equipment is usually related to the material handling system involved. The most commonly applied magnetic separators include pulleys of the permanent magnetic and electromagnetic types applied in belt conveyor systems as head pulleys, suspended magnets applied over belt conveyor chutes or feeders, magnetic drums usually applied as a separately installed unit, plate magnets applied in or above chutes, and grate magnets applied in free-flowing product streams.

For magnetic particle separation and product purification, a selected plant area is determined and the feed is usually brought to that area for treatment. Magnetic separators applied include magnetic pulley separators, magnetic drums of several types, high intensity induced-roll magnetic separators, high intensity cross-belt magnetic separators, magnetic filters, wet high intensity magnetic separators (WHIMS), and special magnet types.

Eddy-current separation is usually applied to the removal of nonferrous metallics from product streams containing nonmetallic or nonmagnetic materials. These include applications such as material recovery facilities (MRF), eg, aluminum can recovery from commingled containers; both preburned

and ash municipal solid waste plants; and nonferrous recovery from scrap fragmentizing operations.

Tramp Iron Magnetic Separation

Tramp iron magnetic separators are used to protect handling and processing equipment such as crushers, pulverizers, and material handling equipment. Separators are usually applied to dry material or material that contains only surface moisture. Iron coarser than 13 mm is usually defined as tramp iron. The size and shape of the tramp iron, together with the material handling system in place or proposed, must be considered when selecting the protective magnet. Magnetic equipment developed for tramp iron removal may involve magnetic head pulleys, suspended magnets, magnetic drums, plate, or grate magnets.

Magnetic Pulleys. An easy and simple way to achieve tramp iron removal from material handled on a conveyor belt is by means of a magnetic head pulley (see CONVEYING). Magnetic pulleys are available in both permanent and electromagnet construction. Most units are of the permanent type, although electrunits are also used for discriminatory separation where field control is required. Magnetic pulleys, relatively low in initial cost and easy to install, accomplish both continuous and automatic trap iron removal. A typical installation is shown in Figure 1.

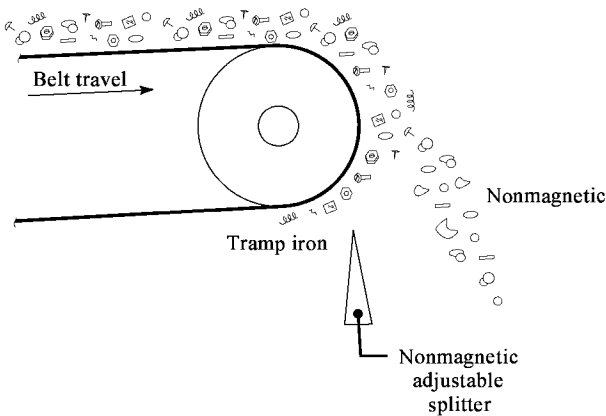


Fig. 1. Principle of operation for magnetic pulleys.

Magnetic pulleys are available in diameters from 152 to 1524 mm and in widths that match the conveyor belt width. The material burden on the belt, belt speed, and type of tramp iron expected to be encountered are all considerations in magnetic pulley selections.

Selection. The magnetic pulley width should match that of the belt. The face width is normally 51 mm wider than the belt width up to 1067 mm wide, and 76 mm on widths in excess of 1219 mm. The speed of operation of the conveyor belt should be determined by calculating the maximum capacity to be handled. Using this information, the diameter of the pulley required to handle the capacity can be determined (Table 2). The operating belt speed must be acceptable for the diameter selected. If the recommended belt speed is exceeded, the pulley diameter that can handle the belt speed must be used.

Inclined belts provide additional areas of contact with the pulley magnetic field and can tolerate higher capacities than normal for a magnetic pulley application. Table 3, which indicates the correction factor to be applied to the capacity, can be used with Table 2 to make the magnetic pulley diameter selection for an inclined belt. The initial cost of units can be related to the capacity or volume throughput. Costs run about \$35 for each m³ h on the smaller units; about \$50 for each m³ h on larger units.

Table 2. Capacities of Magnetic Pulleys for Tramp Iron Removal^a, m³/h

Pulley dia, mm	Width of belt, mm								Normal belt speed, m s
	305	457	610	762	914	1067	1219	1372	1524
305	30	50	100						
381	40	60	115	180					
457	45	80	130	200	300	420			
508		90	140	220	330	470			
610		110	160	260	380	540	750		
762			190	290	440	620	860		
914			210	330	480	690	950	1230	
1067				360	530	760	1040	1350	1670
1219				390	580	820	1130	1460	1820
1372					630	900	1240	1600	1990
1524						950	1300	1680	2090

^a Normal conveyor operation.

Table 3. Correction Factors for Inclined Conveyors

Parameter	Values							
incline, deg	5	6	7	8	9	10	11	12
correction factor	0.955	0.946	0.937	0.928	0.919	0.910	0.901	0.892
incline, deg	13	14	15	16	17	18	19	20
correction factor	0.883	0.874	0.865	0.856	0.847	0.838	0.829	0.820

Suspended Magnets. Suspended magnets having either square- or rectangular-shaped base area are available. The square magnet is most commonly applied and permits installation of self-cleaning additions where such construction is desired. Suspended magnets can be installed at many points in a material handling system, at any point along a conveyor belt, at the discharge end of feeders or screens, and above chutes or launders. Suspended magnets are available in both permanent and electromagnet construction. The permanent type is largely limited to the lighter burden and lower suspension applications. Electro suspended magnets having deep magnetic fields are required when the burden on the belt exceeds the limits of a magnetic pulley.

The burden depth, belt speed, clearance required over the burden, and the size, shape, and weight of the tramp iron all determine the selection and size of the suspended magnet required. The depth of magnetic field produced using a suspended magnet is largely related to the dimensional configurations of the magnet. There is no standardization of magnet lengths or magnet widths. Each manufacturer has design parameters that influence the magnetic field obtained at a specific distance from the magnet face.

The preferred location for a suspended magnet is at an angle over the discharge of a conveyor. At this point the material moves into the magnet face and the load breaks open, making tramp iron removal easier. Suspended magnets can be installed at a variety of other points, but magnet selection must be modified for increased difficulty of tramp iron removal, owing to conditions such as ore on top of the tramp iron or a required change in direction of movement of the tramp iron.

A suspended magnet can be made continuous in the discharge of tramp iron by placing a belt over the magnet face and driving the belt across it. This discharging feature is particularly effective where long pieces of tramp iron are encountered.

Automatic discharging suspended magnets can be operated transversely (Fig. 2) or parallel to conveyor flow (Fig. 3b and 3c). Because the tramp iron must be attracted from a buried location and turned 90° from the movement of the conveyor belt, a larger and stronger magnet is required for a transverse installation. In some instances, the material handling layout, or desired operation location, dictates the use of the transversely mounted magnet.

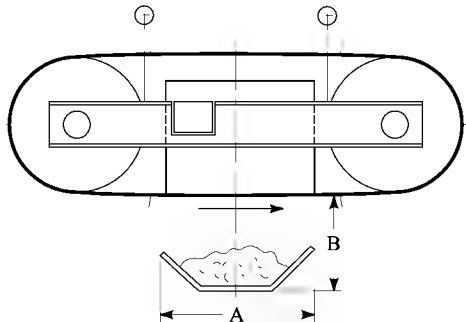


Fig. 2. Self-cleaning suspended magnetic separator over conveyor run (cross-belt), where A = conveyor belt width; B = distance to magnetic belt.

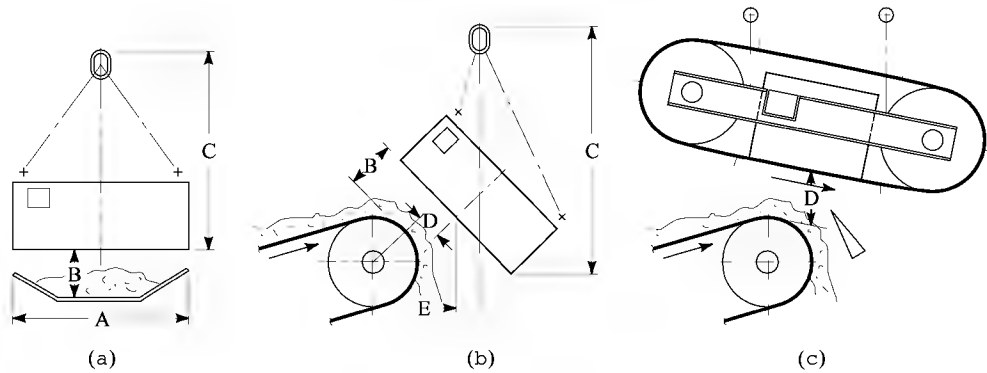


Fig. 3. Suspended magnets, where A = conveyor belt width; B = suspension height; C = overall height; D = distance to magnet centerline; and E = location of lift point. (a) Manually cleaned over conveyor run system, (b) manually cleaned over head pulley, and (c) self-cleaning separator over head pulley (in-line).

Selection. Suspended magnet selection requires determination of the burden depth using one of the following formulas. For installation flat over the belt or transverse self-cleaned mounting, the burden design is calculated as

$$De = \frac{994C}{WV}$$

where De = burden depth in mm, C = capacity in t/h, W = belt width in mm, and V = belt speed in m/s. Capacity for coal is measured in units of 800 kg/m³. For installation at an angle over head pulley or parallel self-cleaned mounting, the burden depth is

$$De = \frac{821C}{WV}$$

For material having a bulk density higher or lower than 800 kg/m³, the resulting De must be multiplied by the factor of 800/ K where K = bulk density of material other than the coal in kg/m³.

A suspension height, Ds , of 75 to 100 mm greater than De , or the maximum lump size, or the edge of the troughed idler in the burden is required, ie, $Ds = De + 75$ to 100 mm.

To determine the size of the magnetic field to be used, the type and size of tramp iron to be removed must be known. A general rule of thumb for 13 to 25 mm balls or cubes is that a 0.1-T (1-kG) field is required at the suspension distance. For tramp iron 50 mm and larger, a 0.05–0.08-T (0.5–0.8-kG) field is required at the suspension distance.

A much more accurate selection procedure is to determine the burden depth, suspension distance, and force index rating of the magnet and particle shape in order to determine the size and strength of the magnet required. From the manufacturer's Tesla curves, the size of magnet required to obtain the required Tesla reading at the determined suspension height can be calculated. Figure 4 shows force index readings for suspended magnets and the relationship to particle shape so that such particles can be picked up from a static bare belt. Force index values to be used in making suspended magnet selections must be obtained from the individual magnet manufacturer and applied as specified. Initial costs vary from \$42 for each m³/h on the smallest units to \$15 for each m³/h on the largest units.

Tramp Iron Magnetic Drums. Magnetic drum separators are used for tramp iron removal where magnetic pulleys and suspended magnets are not feasible. In effect, these are individual pieces of process equipment inserted in the process line. The magnet assembly is held in a fixed position inside the drum shell and the drum shell is driven around this magnet assembly (Fig. 5). The magnet assembly develops a field that typically covers 120–180° of the drum section. These drums can be fed at the top vertical centerline (Fig. 5a) or near the bottom of the drum (Fig. 5b). Overfeed drums produce the highest magnetic removal but result in carryover of some nonmagnetic product. Underfeed drums produce the cleanest magnetic concentrate having reduced nonmagnetic entrapment.

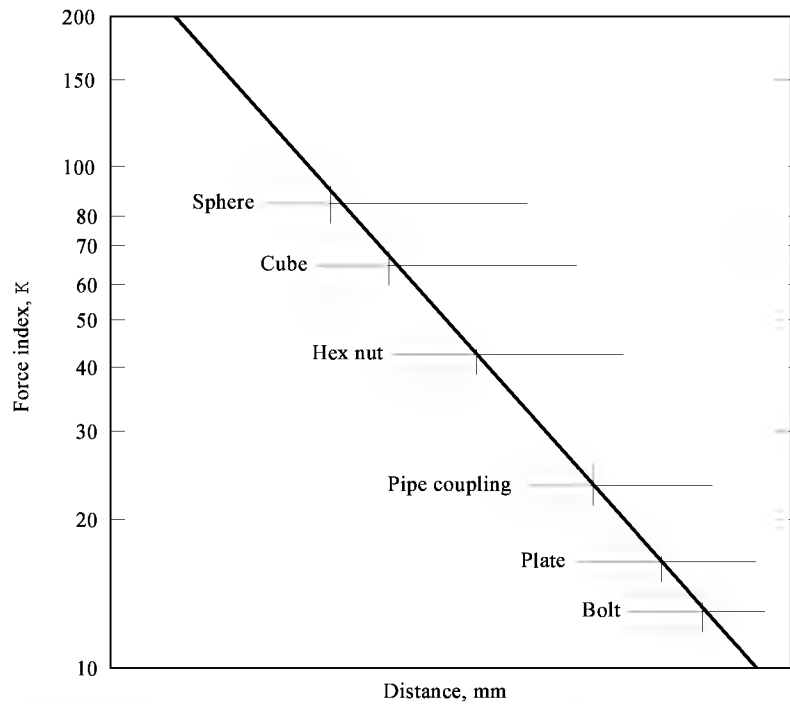


Fig. 4. Static force index (Tesla(ΔTesla Δdistance)) for varying particle shapes, where K is the bulk density of material in kg m^{-3} . The plate has dimensions of $6 \times 76 \times 76 \text{ mm}^3$; the bolt has dimensions of $6 \times 25 \text{ mm}^2$. The horizontal line represents the distance from the magnet face.

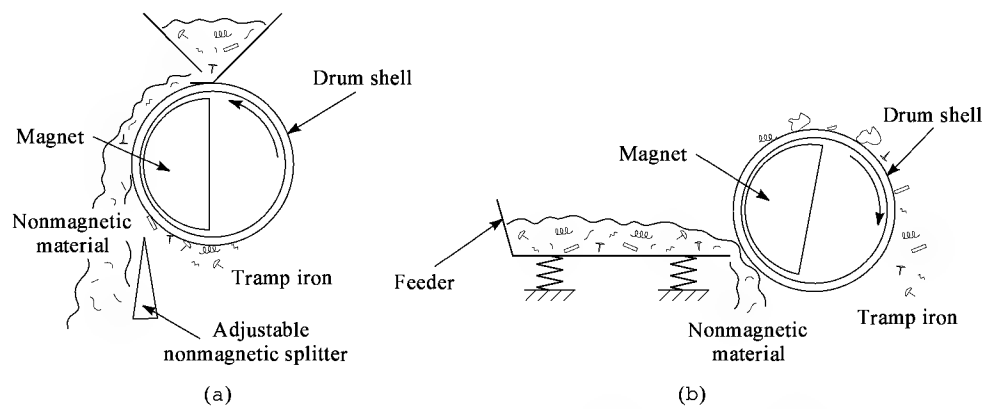


Fig. 5. Arrangements for magnetic drums: (a) overfeed and (b) underfeed.

Magnetic drums are operated either as drums only, ie, on separately mounted framework such as those used in coarse iron cobbing and autofragmentation plants, or as drums enclosed in a housing, such as those used in tramp iron removal from smaller-sized or dusty material. Magnetic drums are selected on the basis of the volume in units of $\text{m}^3 \text{ h}$ to be handled from rated capacities in manufacturers' catalogs. The capital investment for magnetic drums can vary from \$1,500 up to \$81,000, depending on the application and whether an electro or permanent drum is required.

Plate Magnets. Plate magnets (Fig. 6a) are simple devices usually mounted in the bottom of a chute or duct. These magnetized plates are manufactured in various models and are normally of the permanent magnet type. The largest units provide protection to about 115 mm of material depth, at chute angles of up to 45° from the horizontal.

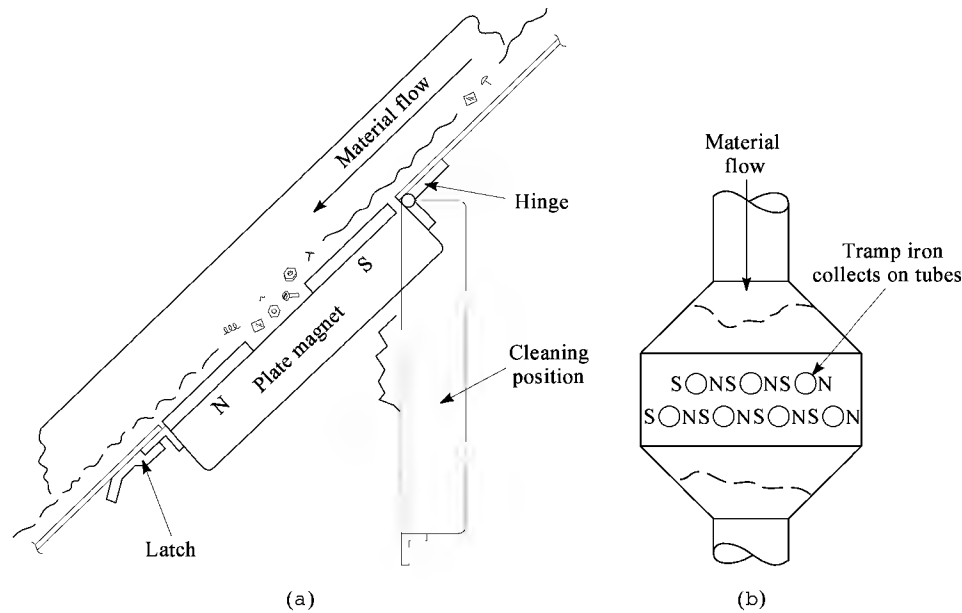


Fig. 6. Schematic of magnetic separation systems where N and S denote the poles of the magnets: (a) plate and (b) grate magnet.

The plate magnet traps the tramp iron against the magnet face so that the trapped material must be periodically removed by hand. Automatic discharging-type plate magnets having timed air cylinder circuits to ensure proper cleaning are available. A chute angle not exceeding 45° is recommended and the plate magnet should be installed as close to the feed point as possible, in an effort to capture the particle before an acceleration increase.

Grate Magnets. The grate magnet (Fig. 6b) consists of a series of magnetized tubes, round or square, mounted in a frame or housing. The collected tramp iron must be removed periodically. Again, periodic self-cleaning mechanisms are available with times cylinders. The grate magnet, largely restricted to use on free-flowing material finer than 13 mm in size, is typically mounted at the discharge of a hopper. Grate magnet selection is made on the basis of volume to be handled, in m³ h, from manufacturers' catalogs. Either style of unit can be supplied with a variety of permanent magnet materials, such as ceramic, Alnico, and rare earth (see MAGNETIC MATERIALS).

Magnetic Concentration and Purification

The magnetic responsiveness of mineral particles provides a positive means of concentrating and or purifying ores. The type of magnetic separator used is influenced by feed condition, ie, wet or dry; the mineral to be concentrated or purified; the relative magnetic responsiveness of the mineral; the feed size; the purity to be obtained, in either the magnetic concentrate or the nonmagnetic product; the capacity to be handled; equipment operating and maintenance costs; and temperature of magnet applications.

Magnetic Separator Classification. Whereas a wide variety of magnetic equipment has evolved over the years, much of this equipment is specialized and has limited usage. A fundamental equipment classification can be made on the basis of feed condition. A wet condition involves the treatment of a slurry or slip; a dry condition involves treatment where the particles are essentially free to move as independent particles. Both conditions depend on liberation of the magnetic particles. If liberation does not occur, the magnet traps the middling particles and reduces the concentration of magnetics.

Wet or dry magnetic separators can be further divided based on the field intensity developed by the individual separator. Broadly speaking, classification by field strength is high, intermediate, or low intensity. Expressed in the magnetic unit Tesla, an arbitrary categorization of these classifications would be >0.5 T, 0.25–0.5 T, and <0.25 T, respectively. Consideration must also be given to the size of the material being treated, the desired purity of either the magnetic or nonmagnetic product, desired capacity, and magnetic recovery.

Commercial Magnetic Separations. Various magnetic separators have been used to accomplish many mineral separations (see MINERAL RECOVERY AND PROCESSING). Some of these include iron ore, ie, magnetite and hematite concentration; abrasive cleaning using silicon carbide and aluminum oxide; garnet concentration; ilmenite concentration; barite purification; bauxite beneficiation; clay mineral cleaning; enamel slip or glaze cleaning; feldspar and kyanite cleaning; magnetite or ferrosilica media recovery in heavy-media separation (HMS) plants; manganese ore concentration; tungsten, monazite, and rare-earth mineral concentration; silica sand cleaning; and columbite and tantalite concentration.

In total numbers of installed units for mineral concentration, the wet magnetic separators generally exceed the number of dry magnetic units. Wet magnetic separators include wet drum separators for media recovery in HMS plants, wet drum separators for iron ore concentration, magnetic filters, and high intensity wet magnetic separators for mineral concentration and clay cleaning. Dry magnetic separators used in mineral concentration include magnetic pulleys for iron ore concentration as well as rare-earth magnetic pulleys for moderately responsive minerals, alternating polarity magnetic drums for iron ore concentration and miscellaneous uses, induced-roll high intensity magnetic separators for silica sand and feldspar cleaning and for ilmenite concentration, and high intensity cross-belt magnetic separators for high value mineral concentration used independently or with induced-roll flow sheets.

Low Intensity Wet Drum Magnetic Separators. Wet drum separators are used to recover ferromagnetic solids from a slurry feed. The principal areas of usage are in media recovery in heavy-media separation plants and in magnetite ore concentration. Physical construction of wet drum separators is slightly different for the two applications. The ore concentrator, which is more rugged in construction, is subject to more detailed specifications than the media recovery units. This is particularly true for feed and collection tank design, bearing construction, wear covers on the drums, magnet assembly design, and magnetic field strength ratings.

Media Recovery Wet Magnetic Drums. In HMS recovery service, the particle size of the feed, particularly the magnetic portion, is quite fine so that drum wear is not a serious problem. Maximum magnetic recovery is important and the highest magnetic purity and solids content in the magnetic concentrate are desired.

Factors to consider in selecting a wet drum separator for media recovery service are the percentage of solids in the feed slurry, the percentage of magnetics in the feed solids, the grade of media used in the operating circuit, the feed volume in m³ h to the separator, the magnetic discharge rate in metric tons per hour (MTPH) from the separators, and the desired magnetic recovery.

Media recovery magnetic drums are conventionally available in 762 or 914 mm diameter and in drum widths to 3.05 meters, usually in increments of 305 mm. The effective magnet width is about 150 mm less than the overall drum width. In some instances 1219 mm dia wet magnetic drums have been used, but the higher capacity of these units does not usually justify the higher costs involved.

Both electro and permanent wet drum separators have been used in media recovery, but permanent-type magnet assemblies are usually preferred because of the savings obtained in the elimination of magnet energization power costs. Both concurrent and countercurrent feed arrangements (Fig. 7) have been used in media recovery plants. The concurrent drum arrangement is usually recommended to give the highest magnetic content in the magnetic discharge and regularly provides the highest percentage of solids in this discharge. The countercurrent unit gives a slightly higher recovery in some applications, particularly when the magnetic loading is heavy.

Feed Solids Content. A good HMS plant operation keeps the medium as free of fines as possible by effective screening of the heavy-media separation vessel feed. Reduced fines reduce viscosity problems in the medium and result in sharper separation of sink and float products. It also improves magnetic recovery on the magnetic drum separators and gives a cleaner magnetic concentrate. The use of cyclones in the HMS circuit, either as the heavy-media separation vessel or as a densifier for rinse or wash water, increases the solids content and must be evaluated in selecting the media recovery wet drum separators for plants in which cyclones are used.

A single wet drum separator is typically used in plants where the feed solids are in the 10–15° range. Extremely dilute feeds make magnetic recovery more difficult and feed volume should be reduced by a factor of 0.5–0.75 for such dilute feed. For solids in the 20–25° range, either single-drum 914-mm units or double-drum 762-mm units are recommended. Above 25° solids, 914 mm dia double drums are recommended.

Magnetic Content of Feed Solids. The magnetic content of the solids is typically >60%. Conventional selection procedures can be used in such cases. When the magnetic content of the solids falls below 60° the feed volume should be reduced and multidrum separators considered in order to maintain higher magnetic recovery.

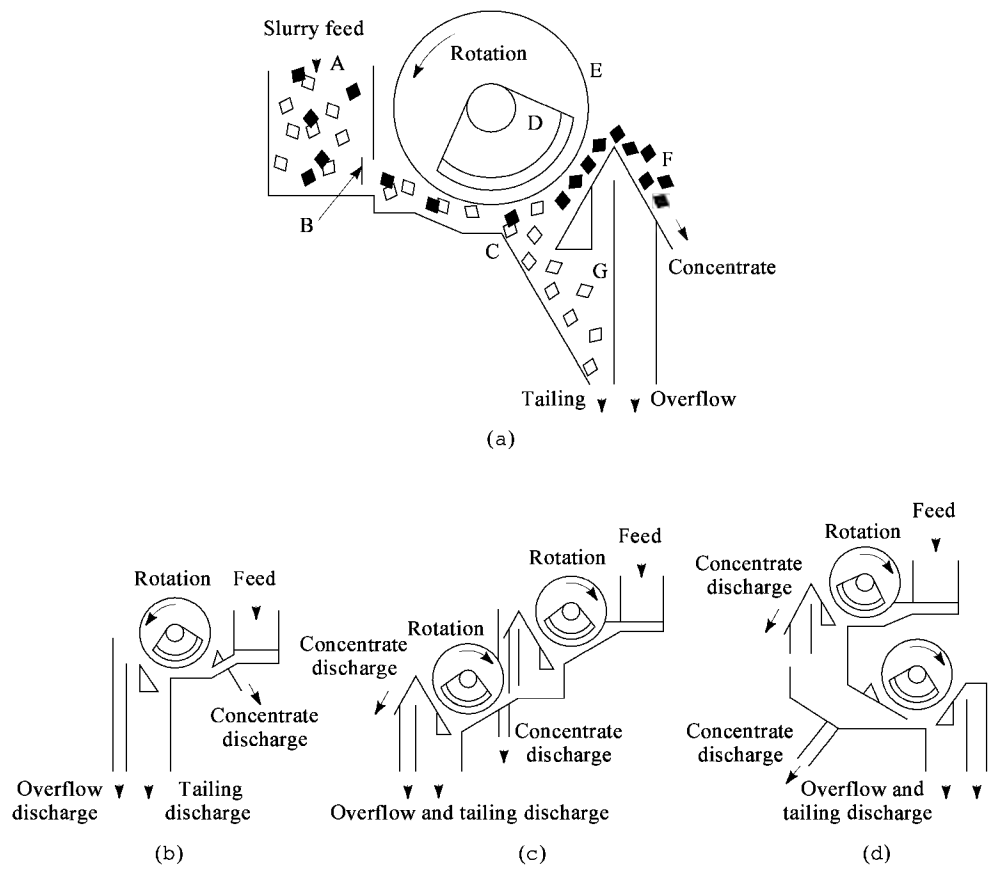


Fig. 7. Magnetic wet drum tank configurations of single and multiple drums: (a) operational features of a concurrent tank, where A represents the slurry distribution box of nonmagnetic stainless steel; B, the tramp screen for oversize material; C, the all-stainless steel tank construction; D, the high strength strontium ferrite magnet assembly; E, the nonmagnetic stainless steel drum shell with O-ring seals for water-tight construction; F, the high density, 60° solid magnetite discharge; and G, the dead box for preventing magnetite accumulation losses; (b) countercurrent tank style; (c) double-drum concurrent concurrent tank style; and (d) double-drum concurrent countercurrent tank style.

Media Size. Media are supplied in several size grades and the grade used varies at each plant. The finer grades improve media stability, but finer particles are more difficult to recover and the feed rate of these finer-grade slurries should be reduced by a factor of 0.5–0.75 to maintain magnetic recovery. A typical size analysis as used in various heavy-media separation plants treating coal (qv) is given in Table 4.

Recommended Feed Volumes. Both feed volume and magnetic solids discharge rate should be balanced in selecting media recovery drum size. The feed rates typically recommended for single and double wet drum separators depend on drum diameter. For a 762 mm dia single drum, the recommended feed rate is 50–55 m³ (h·m); for either a 914 mm dia single drum or a 762 mm dia double drum, 55–70 m³ (h·m); for a 914 mm dia double drum, 70–95 m³ (h·m).

Table 4. Screen Analysis of Ground Magnetite

U.S. Standard sieve ^a		Grade 2, % fine	Grade 3, % medium	Grade 4, % coarse
µm	Mesh size			
+297	+50	trace	trace	8.8
297+213	50+70	3.8	7.1	8.2
213+149	70+100	4.4	8.8	4.3
149+125	100+140	6.2	11.0	14.1
125+74	140+200	7.6	10.0	12.0
74+44	200+325	13.5	14.7	19.9
44	325	64.5	48.4	32.7

^a The + sign indicates retention on the screen; the - sign indicates passage through the screen.

Magnetic Discharge Rate. In order to maintain high magnetic media recovery, the magnetic discharge rates should not exceed 9 MTPH·m for 762 mm dia drums or 15 MTPH·m for 914 mm dia drums.

Wet Drum Operating Considerations. Selection of wet drum separators using the selection criteria discussed produces a media recovery in the 99.6–99.9% efficiency range. Significant operating considerations must be maintained, however, to ensure these recoveries. Some of the most important are as follows: (1) maintenance of a reasonably uniform feed rate and feed distribution across the drum; (2) maintenance of the magnetic discharge point at the proper position. This magnetic discharge point can usually be adjusted slightly, but loose clamp bearings can cause the magnet assembly to fall out of position; (3) maintenance of a proper operating level in the separator tank, ie, loss of level or excessive overflow rates can cause serious loss of magnetic recovery efficiency; and (4) periodic inspection of drums and tanks to determine if holes have developed, which results in leakage into the drum or out of the tank. Wet drum-type magnetic separators for heavy-media separation had an initial cost of about \$8000·m of width in 1995.

Wet Magnetic Drums for Ore Concentration. The depletion of direct shipping-grade ore bodies and the demand for higher grade iron ore concentrates have led to the usage of magnetic drum separators in the concentration of magnetite ores. These relatively low magnetic strength drums can also be used to concentrate other minerals having ferromagnetic properties.

Typically, ore bodies are relatively low in iron content. Iron minerals are finely divided in a gangue matrix. Wet grinding is usually required to liberate the iron minerals, although some beach sands may have liberated iron mineral values. Wet drum separators are limited to the treatment of material <10 mm. The magnetic drum separators applied are usually related to the grinding circuit required to liberate the iron mineral, and are typically designated by application as cobblers, roughers, or finishers.

Cobblers. Magnetic drums used in cobbing services are designated to obtain maximum rejection of a nonmagnetic product and maximum recovery of the iron mineral. Typically, cobblers are applied on a rod mill discharge product. Because the objective is to obtain maximum capacity, these drums are 914 or 1219 mm in diameter and incorporate wear covers on the drum shells to take the wear introduced by the relatively coarse feed size.

Most iron-ore-concentrating drums are applied in a multidrum configuration in order to obtain maximum rejection of nonmagnetic particles. In cobbing service, two drums are typical but as many as four have been used. These drum concentrators incorporate a repulping box ahead of each drum in order to provide the next drum with a feed that can accomplish nonmagnetic rejection.

Laboratory or pilot plant tests are usually conducted on individual ores to determine the number of drums required to obtain optimum concentration results. Drums that are 914 or 1219 mm in diameter are usually used in cobbing service.

Roughers. Magnetic drums used in roughing service, ie, roughers, sometimes called cleaners, are applied after further size reduction has been accomplished on the cobber concentrate. Because the cobber has produced significant upgrading of the ore, the magnetic loadings on these drums is significantly increased. Replaceable drum covers are usually used. The objective is to obtain rejection of nonmagnetics while maintaining magnetic recovery. Typically, a double-drum separator is used in this rougher service, and in some plants countercurrent or combined concurrent and countercurrent drums are used to maintain magnetic recovery under heavy loads. Drums 914 or 1219 mm in diameter are used in roughing applications.

Finishers. Magnetic finishing drums are designed to produce the highest possible iron content in the concentrate. Typically, the feed size has been reduced to a nominal size of 74 μm (200 mesh) or 44 μm (325 mesh) in a ball mill circuit. The feed tank and feed arrangement of the finisher separator is usually of the semicountercurrent design. The objective is to disperse the feed particles in order to obtain maximum rejection of nonmagnetic particles. Both 762 and 914 mm dia drums have been used in finisher applications. Drum covers frequently are not used in finisher construction because of the material size.

Feed Factors in Wet Drum Concentrator Selection. Feed factors in selection of wet drum concentrators include the rate (MTPH), size (10 mm dia max), and volume (m³ h of pulp), as well as the magnetic content, ie, projected magnetic discharge rate (MTPH), and the degree of liberation at each stage, which effects the size distribution. Also important are the desired magnetic recovery and required magnetic field strength. Maximum feed rate varies with the type of separator being considered, ie, cobber, rougher, or finisher. In concentration applications, capacity must be balanced with magnetic recovery and nonmagnetic rejection.

Feed rates for the various types of separators are as follows:

Type	Rate, MTPH m	
	Maximum	Recommended
cobber	75	45
rougher	60	45
finisher	15	15

Owing to the feed pan distance usually maintained on wet drum cobbers, the wear encountered with coarser particles, and the feed velocities required to move coarse particles, the recommended upper size limits for cobber separators is 10 mm in diameter. Individual ore characteristics required to obtain liberation determine the feed size in rougher and finisher feeds. For finishers, where all the nonmagnetics must be overflowed, a sufficiently fine size to accomplish the overflow must be obtained. Typical feed sizes for cobbers are from 841 μm (20 mesh) to 10 mm; for rougher, 420 μm (35 mesh) to 297 μm (48 mesh); and for finishers, 63 μm (270 mesh) to 44 μm (325 mesh). The magnetic content of the iron ores to be concentrated varies over fairly wide limits. Ores as low as 10 wt % Fe have been successfully treated, as have ores having up to 50 wt % or more iron.

Because liberation is not complete at the cobber stage, a substantial amount of attached gangue is carried into the cobber concentrate. Typically, the cobber concentrate contains 50% Fe and accordingly the magnetic content of the rougher feed is quite high. Finisher feed contains 2–5% nonmagnetics and has a high magnetic content, ultimately producing magnetic concentrates near 70% Fe.

Complete liberation of most iron ores is usually finer than 149 μm (100 mesh). Some ores have to be reduced to 44 μm (325 mesh) for successful concentration. In many instances, reduction to a size suitable for pelletizing is required (see SIZE REDUCTION).

Desired Magnetic Recovery. In ore concentration, maximum recovery is desired at all times. Rejection of middling particles, although sometimes desired, is difficult to accomplish on wet magnetic drum separators.

Magnetic Strength Required. There is considerable debate as to the magnetic field pattern required for the various stages of ore concentration. Several styles of magnet assemblies and a variety of pole combinations are available. Average 50-mm readings, ie, average based on readings taken at 50-mm distance and made at the center of each gap and center of each pole, of 0.08–0.12 T are available on 914 and 1219 mm dia wet drum separators. Magnetic drums of this strength have operated successfully in commercial plants. Finisher drums are frequently 762 mm in diameter and develop average 50-mm readings of 0.05 T.

Wet Drum Ore Concentrator Specifications. There is a high degree of individual preference among users for specific construction features in the wet magnetic drums purchased for ore concentration plants. Specifications usually include number of drums, drum diameter, drum shell thickness and material of construction, drum head construction and material of construction, tank type and arrangement, type of magnet assembly, bearing type and lubrication arrangement, type and thickness of wear covers, and effective magnet width.

Drums. Typical selection for the number of drums used in the various applications is two to four for cobbers, one or two for roughers, and two or three for finishers. The inner drum shell, usually 3 mm thick, is specified Series 302 or 304 stainless steel. Drum heads are usually of high tensile strength aluminum alloy or brass. Recessed head bolt construction having an effective seal is specified.

Drum diameters are typically 762, 914, or 1219 mm, all of which have been used in the various basic applications. In some instances, standardization of diameter is specified because of maintenance considerations. Because these separators are constructed more ruggedly and require repulping stages, capital costs are higher than heavy-media separators. Costs in 1995 were approximately \$10,000 m of width for single-drum modules and multiples of the single-drum module.

Tanks. Three basic tank types have been used in ore concentration: concurrent, where feed slurry moves in the same direction as the drum rotates (Fig. 8a); countercurrent, where feed slurry moves in the direction opposite to drum rotation (see Fig. 7b); and semicountercurrent, where there is uprising feed with the drum moving in a selected direction and all tailings rejected through the tank overflow on the opposite side of the magnetic discharge (Fig. 8b).

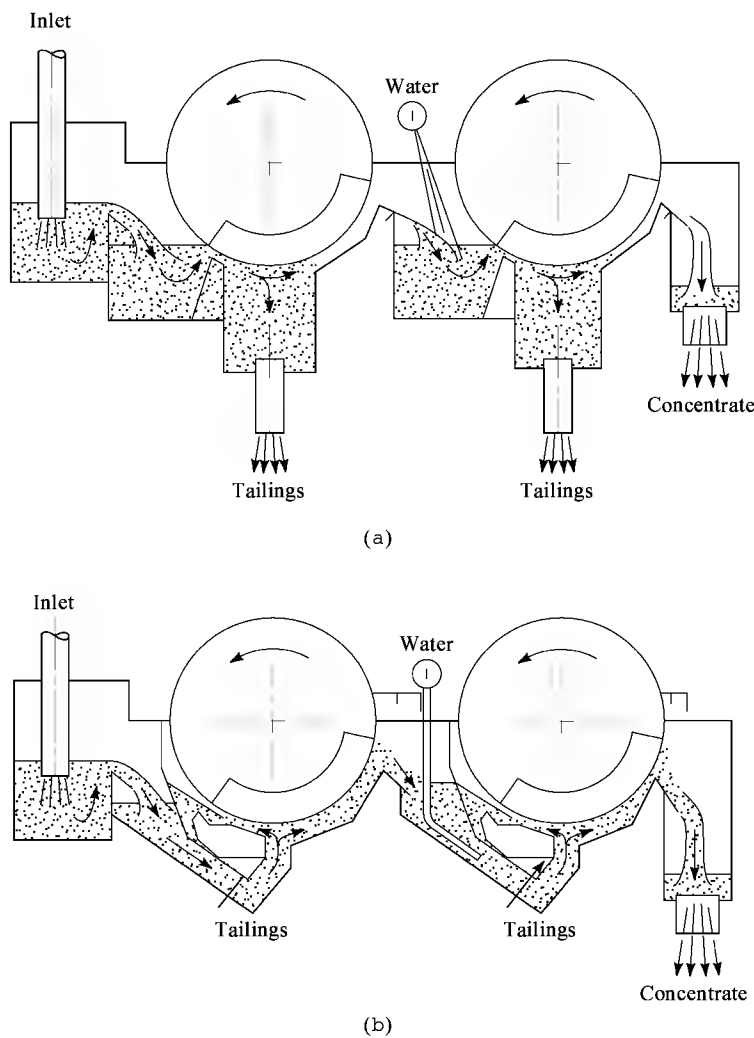


Fig. 8. (a) A double-drum concurrent cobber arrangement; (b) a double-drum countercurrent (Steffenson) finisher arrangement.

An all-stainless steel tank, Series 302 or 304, is frequently specified, but composite tanks of stainless and carbon steel have also been used successfully.

Magnet Assembly. Several types of magnet assemblies and pole combinations have been used successfully. Both electro and permanent magnet assemblies have also been employed. The permanent assembly is frequently preferred because of the power savings effected. Typical pole specifications for the various applications are five or six poles for cobbers as well as for roughers, and from four to ten (normally five) poles for finishers.

The effective magnet width is always less than the total drum width and must be considered when determining the expected capacity of the unit.

Bearings and Wear Shells. Ball-, roller-, and sleeve-type bearings have all been used successfully. Lubrication through the shaft without stopping drum rotation is preferred. A butt of lapped-type wear shell constructed of nonmagnetic stainless steel is preferred for cobber and rougher units. Shells of 3- to 6-mm thickness have been specified, although the shell thickness does influence the surface strength of the magnetic drum and should be taken into consideration when the Tesla pattern is specified. Finishers have been operated using both stainless and soft rubber covers and without any cover at all.

Wet High Intensity Magnetic Separators. There are several types of relatively new, wet high intensity magnetic separators (WHIMS) commercially available. The first units were developed in England for clay purification. This early separator, ie, Jones' separator, was a cycling-type machine that later evolved into the continuous carousel unit (Fig. 9). The Jones' unit has been used to concentrate itabirite in Brazil. These units are capable of developing magnetic field intensities of 2.0 T. Other units have been applied in the purification of silica sand, feldspar, and fluorspar. Concentration of ilmenite, monozite, tourmaline, and chromite has also been achieved by WHIMS separators.

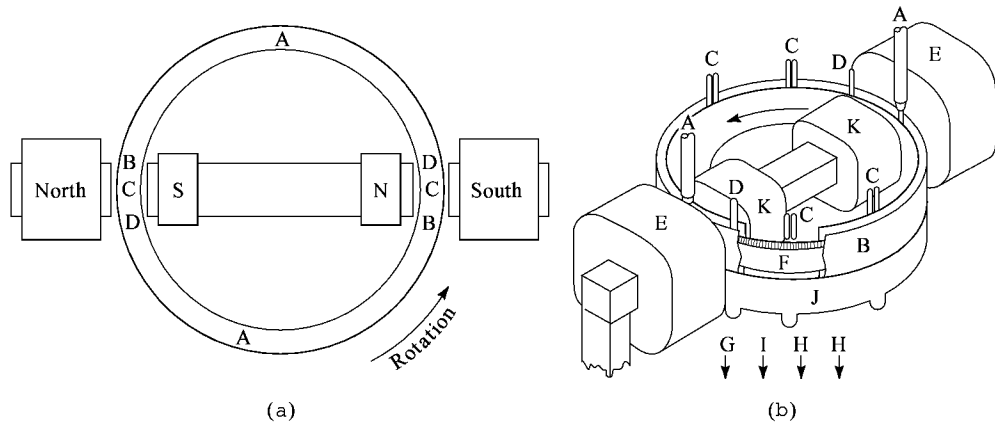


Fig. 9. Continuous carousel wet high intensity magnetic separator where for (a) A is the high pressure scour point (magnetics discharge); B, feed point; C, nonmagnetic discharge; and D, low pressure wash point (middlings discharge); for (b), A is the feed pipe; B, rotor; C, high pressure and D, low pressure water jets; E, outer coil; F, matrix; G, nonmagnetics, H, magnetics, and I, middlings discharge; J, trough; and K, inner coil.

Other wet high intensity units provide configurations that have rotating matrixes similar to wet drum units having cooled electro coils. Still others fall into the category of filters using cryogenically cooled coils and stationary matrixes (Fig. 10).

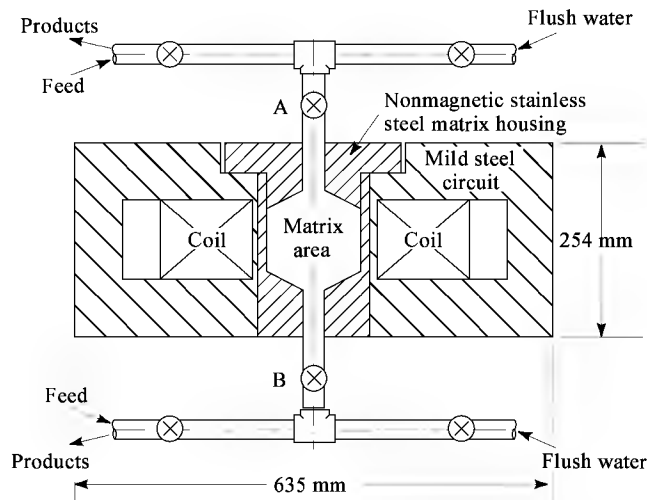


Fig. 10. Wet high intensity magnetic separator using cryogenically cooled coils and a stationary matrix where A is the feed control for top-fed or retention time control for underfed operation and B is the feed control for underfed or retention time control for top-fed operation.

Magnetic Filters. Small magnetic filters are simple devices in which an electrically energized coil or permanent magnets are used to magnetize a magnetic steel grid. The grids develop high points of magnetic strength on their edges. The material to be cleaned is passed through this series of magnetized grids and magnetic particles are attracted to and held onto the grid edges. Periodic cleaning of the magnetic filter is required.

Magnetic filters have been used to clean paint slips and liquids. Filters that are open to the atmosphere or closed, ie, pressure type, are available where the filter inlets are matched to standard pipe connections from 10 to 50 mm, for low capacity applications.

Filter selection, based on the feed capacity, m³ h, to be handled, varies from 1 to 45 m³ h in the many sizes available. A permanent magnetic filter is shown in Figure 11.

Dry Magnetic Separators. Commercial types of dry magnetic separators fall into two broad areas. Low intensity magnetic pulley separators (see Fig. 1) and several types of magnetic drums (Fig. 12) have been used to concentrate iron ore, sponge iron, and to recover iron values from steel mill slags. These separators have effective working magnetic field strength of under 0.1 T and are largely limited to the recovery of ferromagnetic materials. However, rare-earth rolls are available that produce 0.6 T and effectively separate moderately responsive minerals.

High intensity dry magnetic separators include two basic types: the induced-roll and the cross-belt magnet. A high intensity inductively magnetized disk has been used in some areas of the world, but this is not well known in the United States. In addition, high intensity ring-type units have been developed and are used in Australia and Europe, but generally are not used commercially in the United States.

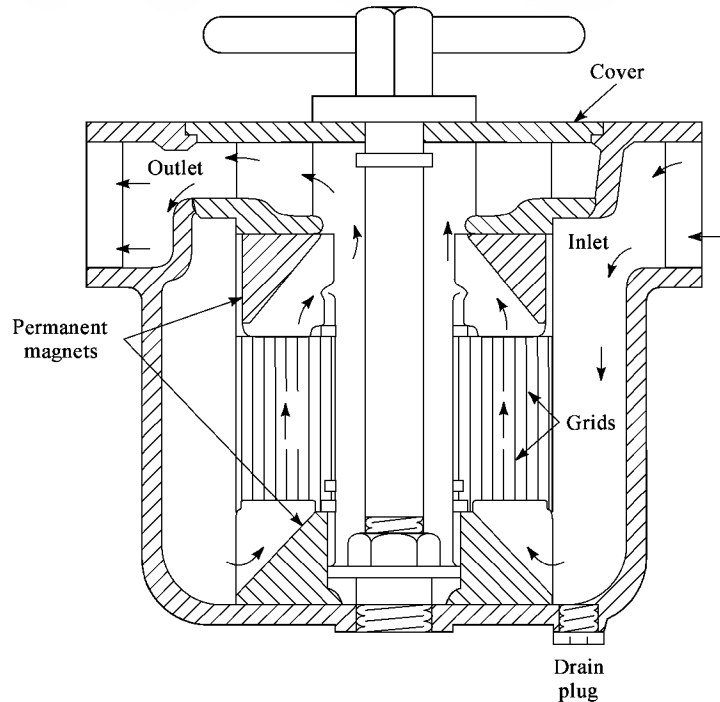


Fig. 11. Magnetic filter.

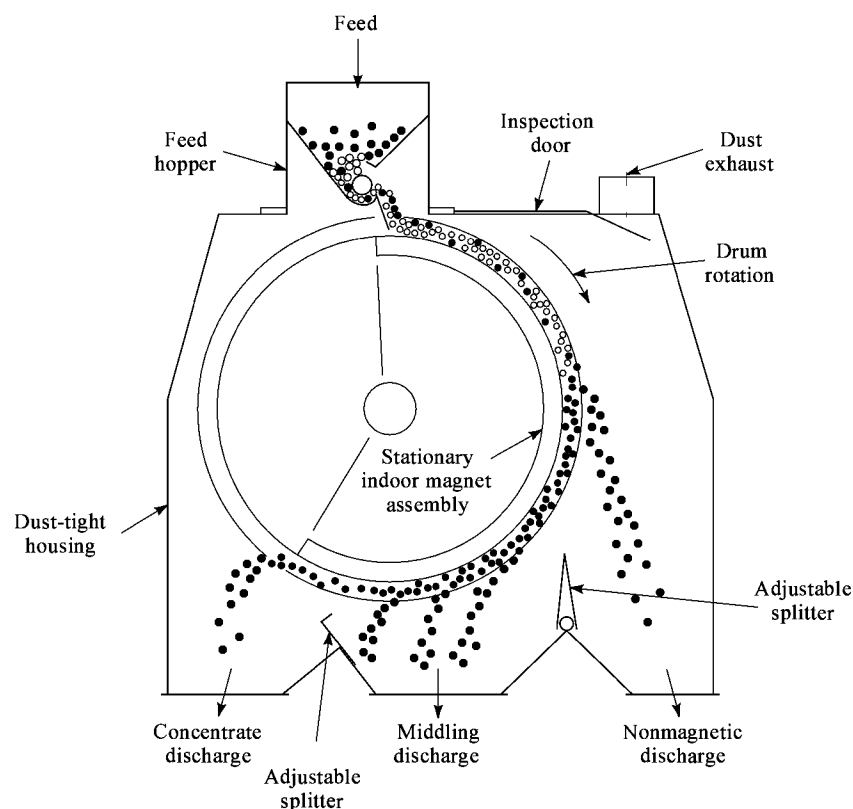


Fig. 12. Magnetic drum for dry feed principle of operation.

Magnetic Pulleys. Magnetic pulleys of special design are used in the concentration of magnetite and other ferromagnetic minerals. For best results, the feed should be screened into various-sized fractions and each fraction treated on a separate pulley separator unit. Typical feed size is 100 × 50 mm², 50 × 25 mm², and 25 × 6 mm². When treating material of 10 mm, an axial pole magnetic pulley should be utilized, as this provides uniformity of field across its width.

The magnetic pulley selection is largely dictated by the tonnage to be handled. The feed should be as uniform as possible across the pulley width. Feed size dictates the diameters of the magnetic pulley used. The larger diameter pulley develops larger magnetic holding capability, as well as a larger radius, which improves holding action of the magnetics on the pulley face. Pulley diameters recommended for various feed sizes are as follows:

Feed size, mm ²	Recommended pulley diameter, mm	Capacity, MTPH m
100 × 50	762–914	240–400
50 × 25	460–610	160–240
25 × 6	305–460	75–135

Magnetic Drums. Two types of magnetic drums, which vary in the construction of the magnet assembly inside the drum, are used in concentration service. The first incorporates a magnet assembly in which the poles vary in polarity across the drum width. This type of drum develops a strong holding force and is used in treatment of ore up to 200 mm in diameter. The second magnetic drum is used for the concentration of finer material, usually 25 mm or smaller in size, and incorporates a magnet assembly of from six to as many as 55 poles that vary in polarity around the circumference of the drum.

Typically, the magnet assembly of the first type, the coarse ore concentrator, has a magnet arc of 120–180°; the assembly of the fine ore concentrator has a magnet arc of 180–240°. The alternating polarity of the fine ore concentrator causes agitation and reorientation of the magnetic particles as the particles traverse the magnet arc, producing the maximum cleaning effect.

Coarse Ore Treatment. Coarse ore concentrating drum selection is based on capacity to be handled and the size of the particle to be treated. As in the case of magnetic pulleys, a sized feed is desirable for optimum operation. Recommended capacity and diameters for coarse ore drums are as follows:

Feed size, mm ²	Recommended drum diameter, mm	Capacity, MTPH m
200 × 150	1524–1829	895–1190
150 × 100	1067–1219	750–900
100 × 50	762–914	240–400
50 × 6	305–460	75–135

Fine Ore Treatment. The alternating polarity drums, used for the concentration of 25-mm ores, are available in both electro and permanent magnet types. The permanent units are usually used on ore finer than 13 mm. The field strength of the electro drum assembly can be regulated by putting a control resistor in the magnet circuit. Control of the permanent drum performance can be regulated to some degree by using a variable-speed drum drive motor to create variations in centrifugal forces exerted on the particles.

A variety of magnet pole configurations are available. The alternating polarity drums are usually available in 460–914 mm diameter and have from seven to 55 poles. Multidrum treatment is frequently used to obtain both additional magnetic recovery and improved magnetic cleaning. The physical arrangement of multipole drums depends on the objective of the separation. The primary magnetic concentrate is retreated on the secondary drum when maximum cleaning is required; the primary nonmagnetic product is retreated in the primary drum when maximum recovery is required (see Fig. 12).

Magnetic drum selection of alternating polarity drum separators is based on the capacity to be handled and the particle feed size. Recommended capacity and drum-type selections for permanent magnets are as follows, where 0 indicates fines:

Feed size, mm ²	Drum diameter, mm	Number of poles	Capacity, MTPH m
----------------------------	-------------------	-----------------	------------------

25 × 13	610–762	6–8	75–90
13 × 6	762	8–30	45–75
6 × 0	914	10–55	15–60

The drums for a feed size of 25 × 13 mm² may consist of either electro or permanent magnets.

High Intensity Induced-Roll Magnetic Separators. Induced-roll magnetic separators are typically used for cleaning such materials as silica sand and feldspar. Both overfeed and underfeed roll designs have been built. The latter is used to obtain a higher degree of cleaning at a relatively high capacity. Owing to the narrow magnetic operating gap, the feed size to induced-roll separators is limited to material finer than 3-mm size. Furthermore, because of the surface activity of very fine material, there should be little 74-μm (200-mesh) material present, unless the loss of this material in the magnetic concentrate can be tolerated.

Induced-roll separators have also been used in the concentration and cleaning of heavy minerals found in beach sands. Examples are the rutile and ilmenite beach sands of Florida and New Jersey. Induced-roll separators are frequently used in combination with high tension or electrostatic separators.

For selective removal of minerals that vary in magnetic response, or for maximum removal of contaminants in silica sand or feldspars, multistage induced-roll separators are applied. Induced rolls from a single roll to as many as seven rolls arranged in series are available. The roll width is limited to 760 mm, largely because of deflection forces exerted on the induced roll by the strong magnetizing sources utilized. Field strengths up to about 2.0 (20°kG) are developed in the working gap.

Induced-roll magnetic separator selection is influenced by the objective of the separation and the capacity to be handled. There is no standard roll diameter, although 100 mm dia rolls are most frequently used in magnetic cleaning services, such as silica sand cleaning. The principle of operation of the induced-roll magnetic separator is shown in Figure 13. Recommended capacities for 100-mm roll diameters are 2–4 MTPH m of width for cleaning operations, and 1–2 MTPH m of roll width for concentrating service. The number of magnetic fields to be used varies with the application and is largely determined by preliminary testing. For silica sand cleaning operations, a three-roll or a five-roll machine is recommended in most cases. The primary roll of most induced-roll separators usually utilizes leak paths of flux from the high strength energizing coil in order to develop a moderate strength primary roll. This roll removes ferromagnetic particles that tend to accumulate and block the secondary rolls if not removed. These units can have a high initial cost per ton treated. Prices amount to about \$15,000 per roll on a 762-mm wide unit.

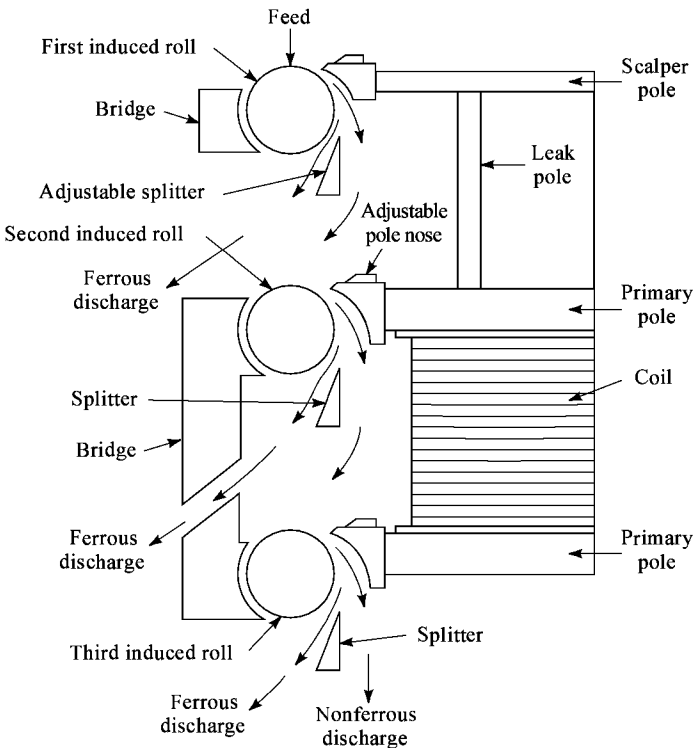


Fig. 13. Induced-roll magnetic separator.

High Intensity Cross-Belt Magnetic Separators. For very selective concentration of weakly magnetic minerals, the high intensity cross-belt separator has been utilized. This separator utilizes a feed belt on which a thin layer of the feed material is introduced to a high intensity magnetic field. The magnetic materials are lifted to the upper pole of this field and transferred by a cross-belt to a collective hopper.

To obtain the high field gradient required for separation of the weakly magnetic minerals, the upper pole is shaped to a point or series of points for improved capacity. The bottom pole is flat. In cases where a variety of weakly magnetic minerals are to be concentrated, a series of high intensity poles is utilized, which have increasing coil strength at each succeeding pole, or by using a variation in the air gap. A high intensity cross-belt magnetic separator is shown in Figure 14.

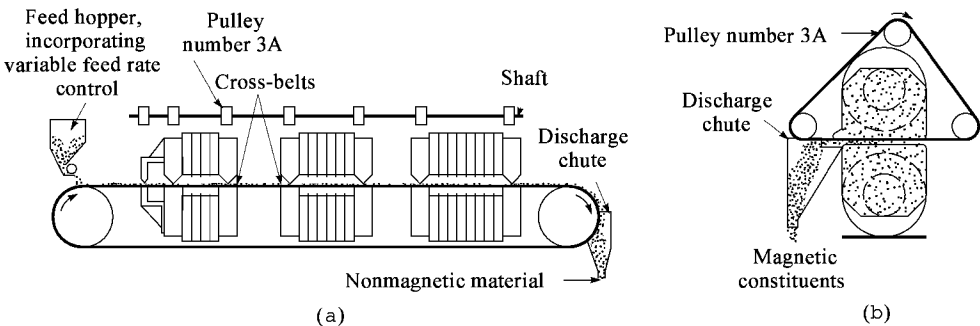


Fig. 14. (a) Schematic of a high intensity cross-belt magnetic separator, and (b) cross-section at pulley number 3A.

The cross-belt separator has been used to concentrate such high value minerals as ilmenite, wolframite, monazite, xenotime, columbite, and tantalite. The use of a lift action provides a high degree of selectivity and a minimum amount of entrapped nonmagnetic minerals. Because the accumulation of magnetics at the high intensity point creates a buildup across the belt width, the cross-belt magnet widths are limited to a maximum of 610 mm. The feed is, at most, a few particles deep and accordingly feed rates are limited to about 2 MTPH m of feed belt width. Lower capacities are required for the cleaning of very weakly magnetic minerals.

A variable-speed drive is usually used on the feed and cross-belt drives to exercise control in separator operation, although the speed is not usually changed once the optimum operating condition is established. Feed rates and the selection of the number of magnetic poles are usually determined by preliminary laboratory tests. The mineral types involved in the feed largely determine the number of poles selected. High intensity cross-belt separators are frequently used in combination with induced-roll or electrostatic separators.

Eddy-Current Separation of Nonferrous Metallics. The advent of material recovery facilities (MRFs) has increased the need for automatic removal of metallic values tenfold. Although many small facilities rely on hand sorting of commingled refuse, ie, glass, aluminum, tin, and plastic, as these plants grow in size, ways of reducing labor costs become more important. Suspended magnets are used to remove ferrous material. The removal of aluminum is relegated to a form of eddy-current separation. A typical processing system is illustrated in Figure 15.

The eddy-current concept was the basis for a separator patented by Thomas Edison in 1889 (1). Only since the mid-1980s has this type of separator been utilized, however, because of the growth in the recycling (qv) and scrap industries. Although the recycling area is of main consideration, there are other areas of usage. These separators are used both in foundries and in autoshrredder operations for the removal of nonferrous metallics from foundry sand and from shredder fluff, respectively. In addition, when applied to aluminum dross (slag), these separators effect the recovery of nonferrous metallics after crushing and sizing; when applied to electronic scrap, they effect the recovery of nonferrous values from shredded circuit boards.

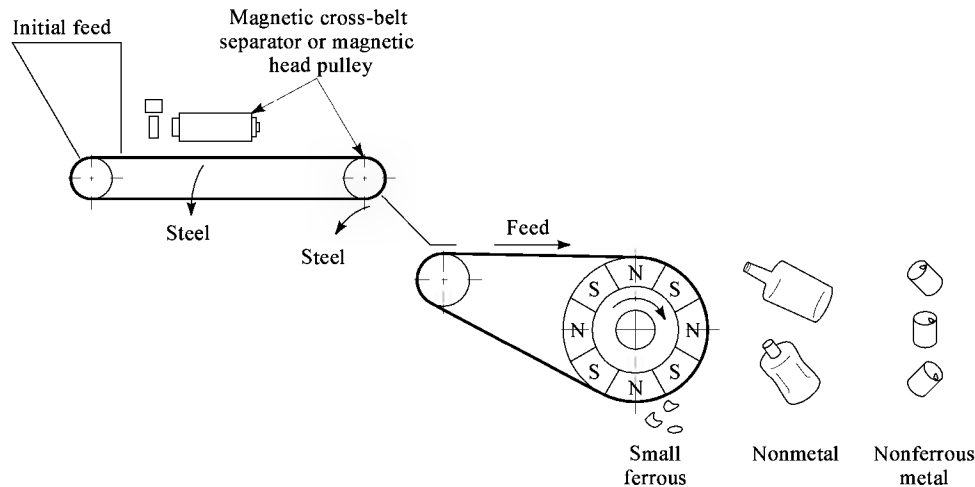


Fig. 15. Schematic of a processing system for a recycling operation.

Principle of Operation. Electrical current flows are induced in all conductors when exposed to an a-c field. These currents generate a magnetic field surrounding the conductors which oppose the field being produced by the a-c field with a force sufficient to repel the conductor. Figure 16 illustrates this principle by showing a rotor consisting of many poles.

The force exerted by a machine is proportional to the field intensity, H^2 , and the frequency. In a rotating device, the frequency is proportional to the number of poles and the rotor rpm. For example, a unit having 10 poles and moving at 3000 rpm produces 15 kilocycles. With the advent of improved rare-earth magnet material, the field intensity for this type of rotor configuration can be greatly increased compared to other types of magnet material, such as ceramic, Alnico, and the early rare earth. Energy products that are nearly seven times greater than those employing a ceramic rotor can be attained.

On a given metallic particle, the repulsive force, F , is dependent on particle mass, M ; electrical conductivity, σ ; density, ρ ; and shape, s .

$$F \approx \frac{M \sigma}{\rho s}$$

The force is proportional to the ratio of a particle's conductivity to its density. Table 5 lists these ratios for several metals.

It is evident from Table 5 that if there were two particles of the same shape and mass, one aluminum, the other copper, the aluminum would have about twice the reaction in repulsive force than would the copper. Both metals have high conductivity, but the densities are nearly three to one.

Particle shape is also important. Disk-shaped as well as cylindrical-shaped conductors have a high response because large induced current loops are formed. Small randomly shaped conductors, such as those present in crushed slag, also respond favorably. Sphere-shaped particles generate small-current loops, however, and do not have a high response. Multiple-current loops occur in conductors that have irregular bends, producing counteractive forces that tend to nullify each other.

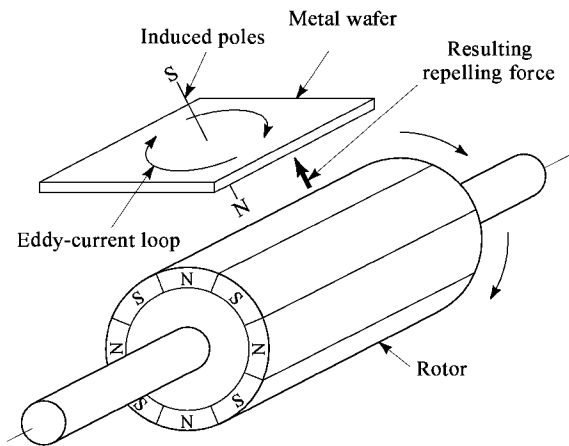


Fig. 16. Eddy-current separation theory of operation.

Table 5. Conductivity–Density Ratio for Metals

Metal	$\zeta \rho^a$
aluminum	14.0
magnesium	12.9
copper	6.7
silver	6.0
zinc	2.4
gold	2.2
brass	1.7
nickel	1.4
tin	1.2
lead	0.4

^a σ conductivity; ρ density.

Manufacture

Magnetic separators are produced worldwide. Some of the principal manufacturers of equipment are as follows:

Manufacturer	Location
Boxmag Rapid	Birmingham, U.K.
Carpco Magnetics	Jacksonville, Fla.
Dings Magnetics	Milwaukee, Wis.
Eriez Magnetics	Erie, Pa.
S. J. Frantz Co.	Trenton, N.J.
Humboldt Wedag	Bochum, Germany
Krupp Industries GmbH	Rheinhausen, Germany
Master Magnets	Birmingham, U.K.
O. S. Walker Co.	Worcester, Mass.
Readings & Lismore Pty., Ltd.	New South Wales, Australia
Stearns Magnetics	Maple Heights, Ohio
Zelba	Spisska Novaves, Czechoslovakia

BIBLIOGRAPHY

"Magnetic Separation" in *ECT* 1st ed., Vol. 8, pp. 624–638, by P. H. Stearns, Stearns Magnetic, Inc.; in *ECT* 2nd ed., Vol. 12, pp. 782–800, by W. J. Bronkala, Indiana General Corp.; in *ECT* 3rd ed., Vol. 14, pp. 708–732, by J. Oberteuffer and I. Wechsler, Sala Magnetics, Inc.

1. U.S. Pat. **400,317** (Mar. 1889), T. Edison.

General References

J. D. Kraus, *Electromagnetics*, McGraw-Hill Book Co., Inc., New York, 1953.

D. M. Hopstock, *Trans. SEM-AIME*, **258**, 222–227 (1975).

Permanent Magnet Handbook, Crucible Steel Company of America.

R. L. Sanford, *Permanent Magnets*, National Bureau of Standards C448, 1944.

J. E. Lawver and D. M. Hopstock, *Mineral Sci. Eng.* **6**(3), 154–172 (1974).

R. S. Dean and C. W. Davis, *U.S. Bureau Mines Bulletin*, **425** (1942).

A. Nussbaum, *Electronic and Magnetic Behavior of Materials*, Prentice-Hall, Inc., Englewood Cliffs, N.J., 1967.

V. G. Derkatsch, *Die Magnetische Aufbereitung Schwachmagnetischer Erze*, VEB Verlag, Leipzig, Germany, 1960.

J. A. Oberteuffer, *IEEE Trans. Magnetics*, **10**(2), 223–238 (1974).

P. W. Selwood, *Magnetochemistry*, 2nd ed., Interscience Publishers, New York, 1956.

J. E. Lawver and D. M. Hopstock, *SME Mineral Processing Handbook*, Society of Mining Engineers, New York, 1985, section 6.

I. S. Wells, *Chem. Eng.* (1982).

G. Clark, *Ind. Minerals*, (May 1985).

J. E. Forciea, L. G. Hendrickson, and O. E. Palasvirta, *Mining Eng.* **10**, 339–349 (Dec. 1958).

J. E. Forciea and R. W. Salmi, *Trans. SME-AIME*, **232**, 339–345 (1965).

B. Skold, *Proceedings of Mineral Processing Meeting of Swedish Mining Society*, Skellerftea, 1970.

G. H. Jones, *Proceedings of International Mineral Processing Congress*, 1960.

J. Iannicelli, *Clays Clay Mineral*, **24**, 64–68 (1976).

E. J. Tenpas, *Rock Products*, (Apr. 1971).

E. Laurilla, *An Approach to the Theoretical Treatment of Magnetic Concentration*, Series A, Vol. 6, Physica-Helsinki, 1958.

H. W. Buus, *Foundry*, (Dec. 1960).

R. K. Singhal, *Mining*, **113**(5), 356–365 (Nov. 1965).

Roberts and co-workers, *Chem. Eng.* 52–68 (June 29, 1970).

W. J. Bronkala, *Chem. Eng.* 113–138 (Mar. 14, 1988).

P. B. Sherwood, *Proceedings of N. S. A. Operators Meeting*, Salt Lake City, Utah, 1976.

R. B. Jacob and J. A. Selvaggi, *Proceedings of AIME Annual Meeting*, Atlanta, Ga., 1983.

H. Harley-Smith, *Magnetic Separation in Mineral Beneficiation*, Metals and Minerals International.

W. H. Goldberger and L. Robbins, *Perry's Chemical Handbook*, 6th ed., chapt. 21.

M. Gupta, *Proceedings of Australian Institute of Mining and Metallurgy*, 1980.

G. Clary, *Ind. Minerals*, (May 1985).

L. A. Roc, *Proceedings of AIME Annual Meeting*, New York, 1958.

R. F. Merwin, *Proceedings of AIME Annual Meeting*, Tampa, Fla., 1966.

C. K. McArthur and P. R. Porath, *Mining Congress*, 28–33 (Oct. 1965).

J. A. Oberteuffer, *Magnetic Separation: A Review of Principles, Devices and Applications*, Institute of Electrical and Electronics Engineers, Inc., 1977.
P. A. Chermemnykh and I. V. Kurchatov, *IEEE Trans. Magnetics*, **28**(1), 651–658 (Jan. 1992).
S. Gröndorfer and co-workers, *IEEE Trans. Magnetics*, **28**(1), 663–667 (Jan. 1992).
D. Fletcher, R. Gerber, and T. Moore, *IEEE Trans. Magnetics*, **31**(6), 4184–4186 (Nov. 1995).

Don Morgan
O. S. Walker Company

SIZE SEPARATION

Size separation, the parceling of particulate material on the basis of size, is an important industrial unit operation that produces, on a continuous basis, coarser and finer streams from a feed stream in a single stage or multiple stages. Multiple stages of size-separation devices can be arranged to produce multiple streams of differing degrees of fineness, such as various grades of aggregates or coals. Multiple staging, ie, multiple stages combined with blending, is also practiced to produce two streams, in the same manner as a single-stage separation, except more efficiently. Examples are the degritting of starch (qv) or paper (qv) pulp (qv). Size separation devices are also used in conjunction with other unit operations, such as closed-circuit crushing and grinding.

Devices for size separation fall into two general categories: those basically involving the probability of passing through an aperture and those that separate by forces of fluid dynamics. The former, probability size separation, is based on the repeated presentations of particles to uniformly sized apertures in screen decks. The latter, fluid-dynamic size separation, takes advantage of the interplay between gravity and drag forces. If the particles are conveyed in a liquid, the separation is termed wet sizing; if conveyed in a gas, the separation is termed dry sizing.

A method of presenting particle size data is as the cumulative fraction (or percentage) finer than (or coarser than) some size, termed a particle size distribution (PSD). The precise definition of particle size is, however, very difficult (1). It is usually defined on the basis of the size analysis method. Fluid-dynamic partitioning is also influenced by other particle properties such as relative density or shape, as is probability partitioning, though to a much lesser degree. Size separation implies that these properties are relatively constant among the particles, and that the size is the basis of the partitioning. If, on the other hand, there are several properties that influence the separation, then this is sorting and interpretation of the separation is not possible (see MINERAL RECOVERY AND PROCESSING). In order to interpret the size separation, the particles must be grouped into subpopulations of common properties.

Screening devices are used to make coarser separations; that is, fine products having 95° passing ca 100-mm–50-μm size. Dry-screening devices have a lower recommended size of ca 500 μm. Wet-screening devices that produce 95° passing ca 500–50-μm size are continually being improved.

Fluid-dynamic separating devices, called classifiers, are used to make finer separations; that is, fine products having 95° passing ca 1000–5-μm size. However, different devices have different separating ranges. For example, hydraulic-settling classifiers based on gravity sedimentation principles produce fine streams in the range of 95° passing ca 1000–100-μm size. Hydraulic-centrifugal classifiers (hydrocyclones) based on multiple *g*-force sedimentation principles, produce fine streams in the range of 95° passing ca 500–5-μm size.

Generally, as the product size becomes finer, the capacity of the separating device decreases. Thus, there are devices that can be fed hundreds of metric tons per hour (MTPH) and produce a 95° passing 50-μm product; but a device that produces 95° passing 5 μm may have a capacity of ca 1 MTPH or less.

Evaluation of Size Separations

Common separating devices are evaluated as follows. Consider a flat screen plate having square openings of dimension *b* and centers of dimension *c*. Ideally, the chance of a spherical particle having diameter *d* passing through an opening would be zero for all particles of relative size *d* / *b* > 1 and one for all particles of relative size *d* / *b* < 1. A plot of the probability-of-passing vs size (Fig. 1, curve D) is a step function, and the separation size, so-called cut size, is *d* / *b* = 1. A perfect separation is one where all particles of size less than the cut size pass and all particles of size greater than the cut size are retained.

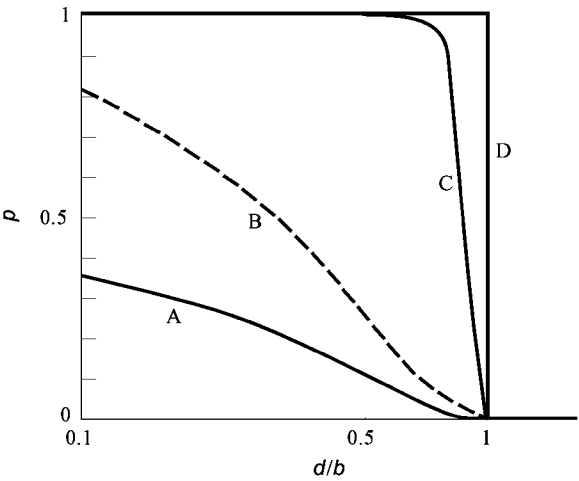


Fig. 1. Probability of a particle size *d* passing through a square aperture of size *b*. See text for discussion of the various curves.

However, the probability of such a particle passing, when approaching the screen plate normally, is as shown in equation 1 (2):

$$p = \left(\frac{b - d}{c} \right)^2$$

(1)

If the *c* / *b* ratio is 1.5, a typical industrial value, then the plot of the probability-of-passing equation vs size (Fig. 1, curve A) is curvilinear. If each particle that did not pass is given another opportunity, then the probability of a particle passing after *n* opportunities is 1 - *qⁿ*, where *q*, the probability of a particle not passing, is equal to 1 - *p*. The plot of probability-of-passing vs size (Fig. 1, curve C) for a very good industrial screen (*n* = 100) is an improved curve approaching the perfect separation step function. Even the curve for *n* = 1000 does not match the ideal curve; rather, it only approaches the ideal because particles of a size close to the dimension of the openings, so-called near size particles, still have a finite probability of not passing.

The characteristic curve is used to evaluate the separation because the deviation from the perfect step function represents a measure of the efficiency of the separation. A commonly used measure to characterize the shape of the curve for a size separation is the sharpness index, given the symbol κ. It is the ratio of the size of a particle having a 75° chance of passing or reporting to the finer stream, ie, *d*₇₅, to the size of a particle having a 25° chance of passing, ie, *d*₂₅. A perfect separation would give a value of 1, whereas merely a splitting action (no size separation) would give a value of 0. The sharpness index of curve C (Fig. 1) is 0.84, 0.92 or 0.9. Another measure not as commonly used is the quartile deviation where one-half the absolute difference

between the upper quartile value, d_{75} , and the lower quartile value, d_{25} , is given the symbol, e_p . The quartile deviation of a perfect size separation would be 0. It is 0.04 for curve C (Fig. 1). Both of these measures represent the inefficiency of the separation.

For size separations, the equiprobable size, ie, the size of a particle that has a 50 : 50 chance of passing, d_{50} , is used to define the cut size. Other cut size definitions for a size separation have been used, including the 95% passing size of the fine stream, $x_{0.95}$; the size at which the cumulative fraction less than in the feed stream is equal to the yield, ie, the ratio of the fine stream flow rate to the feed stream flow rate; and the size at which the cumulative percentage less than in the fine stream is equal to the cumulative percentage greater than in the coarse stream. The d_{50} is gaining wide acceptance as the definition of cut size. The cut size for curve C is 0.9, which is less than the aperture size. This is the reason the aperture size for industrial screens is normally 10–20% larger than the desired cut size.

Curve A exhibits another characteristic: it does not go to one at d/b equal to zero (see eq. 1). This can be viewed as an apparent bypassing, or short-circuiting, of the feed material to the coarse stream, ie, a fraction of all of the particles in the feed did not pass through the screen. If the separating action of a device is viewed as a splitting action (a certain portion of the feed sent to one stream, the rest to a second, independent of size) followed by the size separation of the second stream and the blending of the coarse particles with the first stream, then the apparent bypassing represents the splitting action portion of the separation. This is a second source of inefficiency; hence, the unique value of q at d/b equal to zero is assigned the symbol a . The p values can be corrected for the apparent bypassing by dividing each value by $(1 - a)$. The corrected p values, representing the probability of a particle not split to the coarse stream passing, for curve A are shown (see Fig. 1) as curve B. The κ (0.265) and d_{50} (0.3) values must then be determined from this corrected curve, not the uncorrected one. Apparent bypassing of the feed material to the fine stream is also possible; however, such an occurrence would represent a device malfunction, eg, big holes in the screen, which is correctable. Apparent bypass cannot always be corrected, and can be present even for a corrected curve such as curve C because of the splitting action of the device. Thus a separator can have a constant corrected curve with increasing feed rate, but an increasing a value. If a size separation device is to be utilized for dedusting, ie, the production of a fine particle-free coarse stream, then obviously it should not have an apparent bypass value.

By solving the appropriate convection–diffusion equations (3) for the movement of solids in fluids, either at free-settling or hindered-settling conditions, similar characteristic curves result that increase in steepness (the sharpness index increases) and decrease in the cut size with increasing time. These curves also can have apparent bypassing to the coarse stream, but should not have apparent bypassing to the fine stream. A typical curve, a plot of the probability-of-reporting-to-the-coarse stream vs log size, is shown in Figure 2.

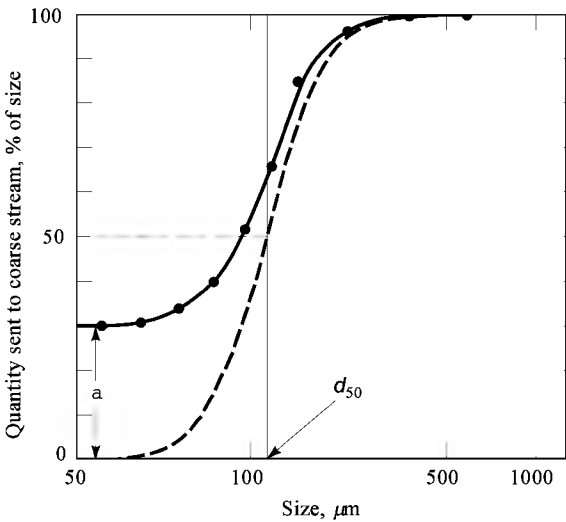


Fig. 2. Probability of a particle reporting to the coarse stream in a fluid dynamic classifier where the vertical line is the ideal classification at d_{50} ; (—) is the separation curve; and (---), the corrected curve.

The characteristic separation curve can be determined for any size separation device by sampling the feed, and coarse and fine streams during steady-state operation. A protocol for determining such selectivity functions has been published (4). This type of testing, when properly conducted, provides the relationships among d_{50} , κ , and a at operating conditions. These three parameters completely describe a size separation device and can be used to predict the size distribution of the fine and coarse streams.

All three parameters, the cut size, sharpness index, and apparent bypass, are used to evaluate a size separation device because these are assumed to be independent of the feed size distribution. Other measures, usually termed efficiencies, are also used to evaluate the separation achieved by a size separation device. Because these measures are dependent on the feed size distribution, they are only useful when making comparisons for similar feeds. All measures reduce to either recovery efficiency, classification efficiency, or quantitative efficiency. Recovery efficiency is the ratio of the amount of material less than the cut size in the fine stream to the amount of material less than the cut size in the feed stream. Classification efficiency is defined as a corrected recovery efficiency, ie, the recovery efficiency minus the ratio of the amount of material greater than the cut size in the fine stream to the amount of material greater than the cut size in the feed stream. Quantitative efficiency is the ratio of the sum of the amount of material less than the cut size in the fine stream plus the amount of material greater than the cut size in the coarse stream, to the sum of the amount of material less than the cut size in the feed stream plus the amount of material greater than the cut size in the feed stream. Thus, if the feed stream analyzes 50% less than the cut size and the fine stream analyzes 95% less than the cut size and the fine stream flow rate is one-half the feed stream flow rate, then the recovery efficiency is 95%, the classification efficiency is 90%, and the quantitative efficiency is 95%.

There are relationships between the independent size separation device parameters and the dependent size separation efficiencies. For example, the apparent bypass value does not affect the size distribution of the fine stream but does affect the circulation ratio, ie, the ratio of the coarse stream flow rate to the fine stream flow rate. The circulation ratio increases as the apparent bypass increases and the sharpness index decreases. Consequently, the yield, the inverse of the circulating load (the ratio of the feed stream flow rate to the fine stream flow rate or the circulation ratio plus one), decreases; hence the efficiencies decrease. For a device having a sharpness index of 1, the recovery efficiency is equal to $(1 - a)$.

There is a unique d_{50} value which produces a fine stream having a specific 95% passing size from a feed stream. The feed size distribution should analyze at least 50% less than the 95% passing size of the product (5). However, the necessary d_{50} value varies with the sharpness index value. In general, the required d_{50} decreases, as the sharpness index decreases (eq. 2):

$$d_{50} \sim (\kappa + 0.16) x_{95} \tag{2}$$

This is undesirable because generally the capacity of a device is much less when operating at a lower d_{50} value.

Screening

Screening is a process whereby particles are presented in an appropriate manner to a series of apertures of fixed dimensions that allow finer particles to pass through into the undersize, and coarser particles to be retained in the oversize. In addition to sizing, screens are used for desliming, ie, the removal of very small particles from much larger ones by draining or rinsing; dewatering (qv), ie, the removal of water from particles by draining; and prewetting, ie, the addition of water to the particles by spraying.

Equipment. Screening devices can have stationary or moving screen decks. Moving screen decks can be designed as rotating cylinders (trommels) or vibrating surfaces. By far, the most popular screen deck design for sizing is the inclined vibrating screen (Fig. 3). Components of vibrating screens are the vibrating frame, deck support frame, screening deck, motor drive assembly, and feed box distributor. Auxiliary components include conveyor belts, feed chutes, and dust collection systems. Vibration is produced on inclined screen decks by circular motion in a vertical plane of 3–12.5-mm amplitude at frequencies of 700–1000 cycles per minute. Newer high speed screening designs operate at 3600 cycles per minute. The vibration lifts the material, producing stratification. Stratification places the larger particles on top, and the smaller particles on the bottom, next to the apertures. This also helps push near-size particles through the apertures, reducing blinding. When the screen deck is used on an incline, the material cascades down the slope, introducing the probability that the particle either passes through the opening or over the screen surface. For horizontal screen decks, the motion must be capable of conveying the material without the assistance of gravity. Straight-line motion at an angle of approximately 45° to the horizontal produces a lifting component for stratification and a conveying component for probability of separation as the material passes across the horizontal screen surface. Industrial screens typically range from 2–7.5 m in length.

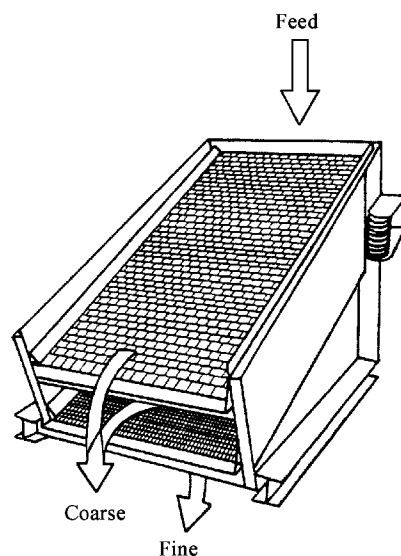


Fig. 3. Inclined vibrating screen.

Some of the numerous factors influencing the passage of a particle through a screen opening are ratio between cross-section of particle and of opening, percentage of screen open area, angle of incidence of feed, efficiency of spread of feed over screen area, kinetic energy of particle approaching screen opening, moisture of feed, stickiness of particle and of aggregated particle, pressure of particles riding above those next to the screen deck, blinding of screen apertures, corrosion of deck material, electrostatic bunching, shape of particle, amount of near-size particles in the feed, rate of feed, thickness of the bed, tautness of deck, shape of screen apertures, trajectory of particle imparted by amplitude and frequency of vibration, bulk density of material, and particle integrity (6). Because of the multitude of factors acting individually and interactively, no fundamental procedure based on first principles has been developed for sizing commercial screens. For example, increases in the quantity of near-size particles, ie, particles where the sizes lie between 0.5 and 1.5 times the aperture size, lead to more blinding, which in turn lead to a decrease in the percentage of open area, ie, the ratio of the aperture area to the screen deck area, which reduces the screening efficiency of the screen deck. Most manufacturers of vibrating inclined screens have developed variations on a design theme that involves estimating the design loading, defined as flow rate / screen area, and then dividing the design feed rate by the design loading to obtain the required deck area (7).

Design Loading. The design loading, or capacity, is estimated by starting from a basic loading for square screen cloth having the desired aperture size. The base loading decreases with decreasing aperture size, partially because the open area decreases. The base value assumes an optimal bed depth, deck slope (flow velocity), quantity of near-size material, free-flowing material, and screen open area. Thus, the base loading must be modified for material properties, equipment design, and operating conditions that differ from the standard ones. For example, there are modifying factors for bulk density, size distribution, and particle shape. The capacity decreases with increasing surface moisture for materials that become cohesive. When the problem is severe, wet screening, ie, water sprays mounted above the screen deck, should be considered. Wet screening increases the capacity, particularly for smaller aperture sizes. The base value is modified for percentage of open area, because capacity decreases directly with decreasing open area, and for aperture geometry, because capacity increases with rectangular apertures, but decreases with circular apertures, because the open area increases and blinding decreases. The base value also depends on whether the screen deck is on the top, middle, or bottom, because larger aperture decks are usually placed above smaller aperture ones in order to protect the smaller aperture deck from damage by large particles. The value is modified for deck slope and motion, ie, frequency, amplitude, and direction.

Deck Area. After the design loading has been estimated, the design feed rate is divided by the design loading to get the effective deck area. This value is usually increased by at least 10% to take into account blockage by hardware associated with the deck support frame. In order to convert the effective deck area into length and width dimensions, the deck width is calculated based on a standard bed depth. The width is selected such that the cross-sectional area of the volume of theoretical oversize leaving the screen deck has a height that is twice the aperture opening. Then the length is calculated by dividing the effective area by the width. Because the length determines the screen efficiency and the width determines capacity, the length should be at least twice the width.

Final screen selection is such that the length and width dimensions match off-the-shelf machines in order to keep costs down. Operating the installed screen at a feed rate greater or less than the design rate results in a less efficient screening operation, because bed depth varies with feed rate.

The top size of the feed to a screen deck should not be greater than two to four times the aperture size of the deck. Double- (see Fig. 3) or triple-deck screen arrangements are used, requiring a separate sizing for each deck. Then the final screen size is set by the largest deck.

Screen decks can be woven wire screens (including piano wire), perforated screen plate, or profile wire. Materials used range from ferrous and nonferrous alloys to rubber and plastic. Screen decks made of rubber or plastic flex reduce blinding, as does unique woven wire geometry. The addition of bouncing balls underneath the deck, or air sprays similar in concept to water sprays over the deck, are also used to reduce blinding. Vibrating the screen deck, at very small amplitude but very high frequency, has also been used successfully to reduce blinding.

Data for dry screening on a 20-mm square aperture vibrating screen (8) indicate that the screen is relatively efficient, giving an apparent bypass value of 0.5%, sharpness index of 0.8, and a cut size of 20 mm. On the other hand, results (9) from a plant operating an 8 ft × 20 ft (2.4 m × 6.1 m) double-deck screen with 16 mm woven wire bottom screen deck gave an apparent bypass of 4%, sharpness index of 0.56, and a cut size of 17 mm. Data (10) for smaller

apertures (0.5 mm) indicate that the cut size is smaller than the opening size, and higher sharpness indexes are only achieved using long screen lengths.

Slurries. Mixed results have been reported (11,12) for screening slurries using small (100- μ m) aperture vibrating screens. The ratio of cut size to aperture size decreases from 0.75 to 0.5 with increasing solids concentration. Because the sharpness index is directly proportional to the ratio, it decreases from 0.7 to 0.5. The apparent bypass value, which is essentially equal to the water split, increases with increasing feed rate and solids concentration. However, there are indications that the type of screen cloth used can affect these results when going from a square to an elongated rectangular aperture.

There is a family of stationary screens known as cross-flow screens, used mainly for dewatering or desliming, that have been applied to slurry sizing. The original member was the sieve bend where feed entering a curved sieve separates into two streams. The fines go through the slotted openings; the coarse stream follows the bend and exits off the end of the sieve. These screens are characterized by a slotted deck, usually made of stainless steel profile wire and set at an angle. The slurry flow is at right angles to the slots and, depending on the angle of inclination, the cut size value is one-half to two-thirds of the slot dimension. Hence, it is possible to size at 0.5 mm using a 1-mm slot, which should reduce blinding problems. However, as the slot dimension is reduced, the apparent bypass value increases, approaching 100% for a slot width of 40 μ m. This can be substantially reduced (values approaching 5%) by vibrating the screen deck (13). Rapping, ie, using a large amplitude and low frequency, is not recommended. A typical sharpness index value is 0.5.

Probability Screens. Other approaches to reducing the blinding effect have included the development of so-called probability screens. Screen blinding led to the commercialization of a simple principle. The projected area of a rectangular aperture, normal to the horizontal, decreases as the angle of inclination of the screen deck containing the aperture increases. Thus, the probability of passage for a particle falling normal to the horizontal, which would normally pass through the aperture, decreases with an increasing angle of inclination. The probability of a particle lodging in the opening is minimal. The commercial sizer (Fig. 4a) consists of five superimposed screen decks, each with a slightly steeper slope and a smaller aperture than the one above. The screen apertures are no less than twice the desired separation size. Two out-of-balance motors transmit a linear vibration to each deck, and the resulting action fluidizes the feed so that the particles are presented essentially normal to the horizontal. This design can achieve separation sizes of 50 μ m down to 100 μ m at high capacity and minimal blinding compared to conventional screens.

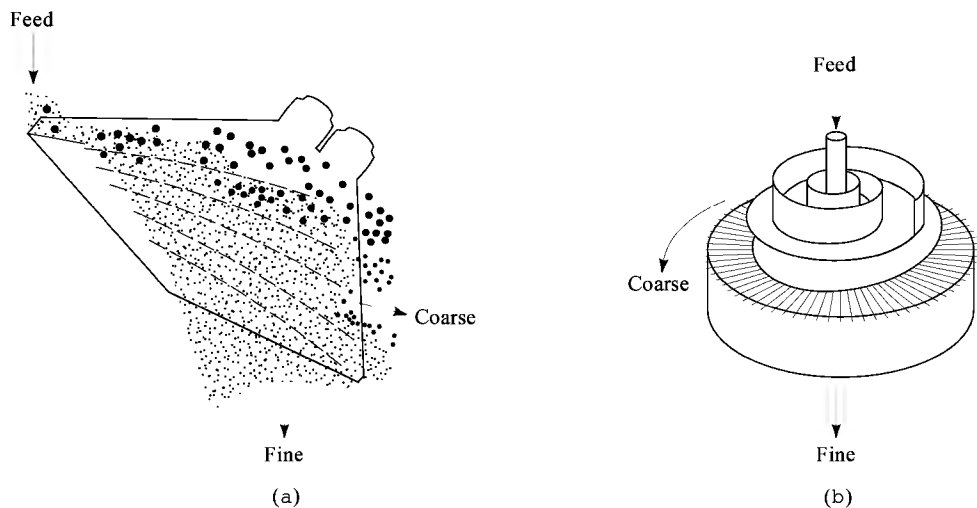


Fig. 4. Probability screens: (a) hi-prob sizer; (b) Ropro screen.

Another screening design (Fig. 4b) developed to handle feed materials with extensive damp, clayey fines content is the rotating probability screen. The Ropro was developed by the National Coal Board in the U.K. (14) after it had rejected electrically heated decks, piano wires, loose rod decks, air sprays, bouncing balls, centrifugal screening, and flip-flop decks made from polyurethane that stretched, relaxed, and fluidized the material to be screened. The screen of the Ropro is created by fitting stainless steel rods to a vertical rotating shaft. The rods, radiating from the central hub, create a horizontal circular surface. The apertures between the radial elements of the screen progressively enlarge and the rods have no supporting members. These apertures are larger than the feed. If the deck is stationary, all the feed material passes through the apertures; if the deck is rotated at high speeds, none of the feed material passes through the apertures. Thus, by controlling the speed of rotation, the screen size is regulated. Hence, the variable rotation speed creates a variable aperture screen. Although no blinding is claimed for this design, no screening efficiency data, particularly apparent bypass, have been reported. An upper deck, designed to scalp the feed to the lower deck, and a rotating table product-collecting system are available modifications.

Wet Classification

Settling-Pool Classifiers. The settling-pool group of fluid-dynamic separating devices (classifiers) (Fig. 5) consists of a rectangular tank having an inclined floor that creates a pool. Feed slurry is introduced at the side of the tank and the overflow of fine particles and water exits through an overflow weir and box arrangement. The coarse particles settle to the bottom and are discharged over the upper edge of the tank after being dragged up the inclined floor by some mechanism, such as intermittently operating rakes, continuously operating drag conveyors, or continuously operating spirals.

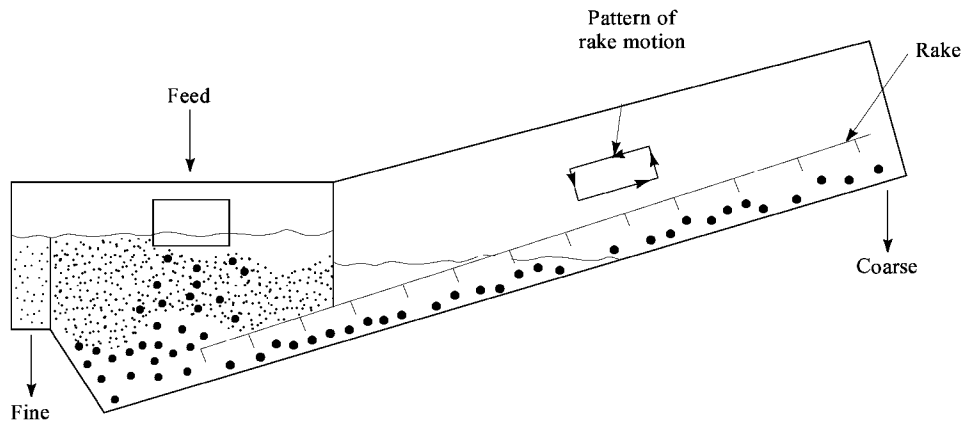


Fig. 5. Settling-pool classifier showing the rake and the pattern of rake motion. Incline is from 14–20° (25–35 cm m).

The length of time the particles stay in the pool determines the distance the particles settle in the pool. Thus, the feed entrance must be located so that the velocity of the pulp toward the weir, together with the distance, allows sufficient time for the fine particles to be carried out over the weir as the

coarse particles settle out below. It is assumed that if a particle settles more than twice the overflow weir crest, it settles out as a coarse particle. The coarse particles are agitated and washed as they are conveyed up the inclined floor, thus reducing the amount of undersize particles carried out in the coarse stream. The pulp velocity, and hence time in the pool, can be varied by altering the weir geometry, the total volume flow, or the solids' concentration. All basically change the mean residence time, ie, the ratio of the mass of holdup material to the feed rate, by varying the mass of holdup material or the feed rate. However, some changes may force the pseudo-free settling conditions more toward hindered settling conditions, thus giving an unexpected result (see SEDIMENTATION).

The separation mechanism of a settling pool should be very straightforward (3). If a cylinder of slurry is shaken initially so that all the particles are uniformly distributed throughout the suspension, then all the particles immediately start to settle under the influence of gravity at rates determined by size, shape, density, and other factors. The upper half of the cylinder then contains particles given to the overflow whereas the lower half contains particles given to the underflow. A certain size of particles just settles from the top to the level at which the overflow and underflow are partitioned. The underflow contains all particles having settling rates greater than the certain size, particles having settling rates less than the certain size because those were near the level between the overflow and underflow, and particles of all settling rates which were originally distributed into the underflow. This last is called void-filling material and represents apparent bypassing for this type of classifier. Dividing the quantity of particles of each size remaining in the upper half by the quantity of particles of the same size in the initial slurry estimates β . The cut size and the apparent bypass values decrease with increasing settling time; the sharpness index increases, becoming essentially constant at about 0.6. The apparent bypass value is approximately equal to the water split, ie, the ratio of the amount of water in the lower half divided by the amount of water in the initial slurry. Superimposing turbulence upon this process produces the same pattern, except that the cut sizes are higher and the sharpness indexes lower after the same settling time.

Analysis of this type of classifier (15) suggests that the sharpness index is between 0.5 and 0.6, consistent with calculated results, because the degree of turbulence can be high in these devices (16). A DSF Dorr classifier (1.8 m $\times$ 7 m), operating at 19 strokes per minute and having a weir depth of 100 cm and a slope of 19.4 cm m, produced a cut size equal to 240 μ m, a sharpness index of 0.5, and an apparent bypass of approximately 26% when the water split was 26% (15).

As the settling type of classifiers are used to produce finer size separations, the pool volume is increased. A settling-type classifier can be operated to produce cut sizes between 150 and 25 μ m but having a corresponding decrease in the sharpness index from 0.5 to 0.2, apparently caused by the increased influence of turbulence (16). Also, the percentage of solids in the fine stream is reduced from 25 to 7% and the coarse stream percentage of water increases from 15 to 25%. This increases the apparent bypass percentage or void-filling material. In addition, because the maximum flow rate of the coarse stream is fixed for each device, the feed rate must be reduced, thereby reducing device capacity.

Hydrocyclones. Increasing the gravitational force by developing centrifugal force decreases the settling time of smaller particles. A centrifugal force can be imposed on particles by rotating the slurry of particles (see SEPARATION, CENTRIFUGAL SEPARATION). The classifying hydrocyclone (Fig. 6) is designed to rotate the slurry of particles by introducing it tangentially into a cylinder, thereby achieving essentially free vortex motion. A second, smaller-diameter (20–40% of the cylinder diameter) pipe, called the vortex finder, is placed through the exact center of the solid top of the cylinder, extending substantially below the feed inlet but above the bottom of the cylinder. The slurry that exits through the vortex finder is termed the overflow and contains the finer particles. Attached to the bottom of the cylinder is an inverted truncated cone section, the included angle of which ranges from 10–30° and is typically 15–17°. The slurry that exits through the cone section (apex) is termed the underflow and contains the coarser particles. The hydrocyclone is the most popular industrial wet classifier.

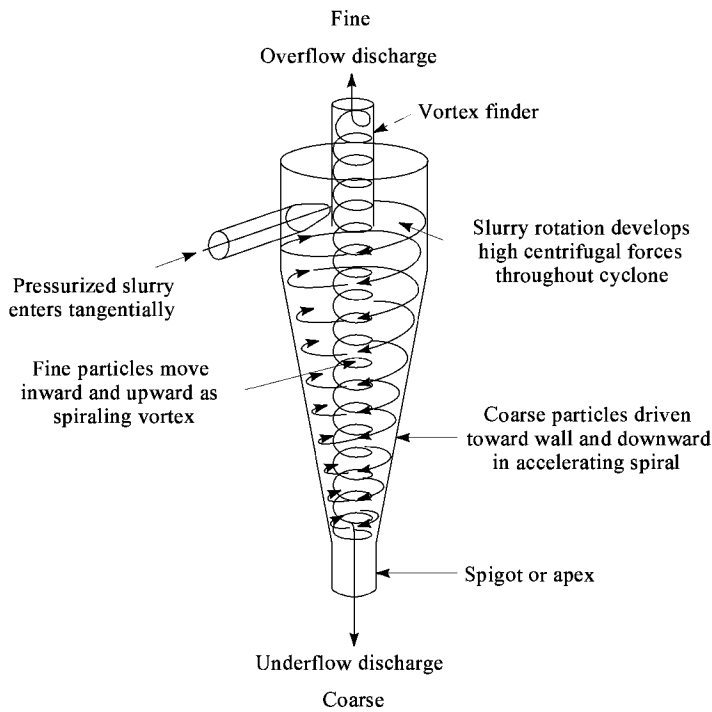


Fig. 6. Hydrocyclone classifier.

As the fluid rotates in the hydrocyclone, forming an air core in the center, there are three fluid velocities of interest: the tangential, the radial, and the axial (17). The tangential velocity, which is zero at the cyclone wall, increases immediately upon moving from the outer wall to the air core, reaching a maximum and then decreasing. There is a locus of constant tangential velocity. The radial velocity immediately increases from zero at the cyclone wall, but then, moving from the outer wall to the air core, decreases to zero. In the cone section, the axial velocity immediately drops from zero at the cyclone wall. A positive value indicates vertical movement. Then, moving from the outer wall to the air core, the axial velocity increases to positive values. Thus, the fluid motion is down the wall of the cyclone to the apex and up the air core through the vortex finder. In the cylindrical section, the axial velocity goes negative again, approaching the vortex-finder wall. The fluid flow is then down the inner cyclone wall and the outer vortex-finder wall. There is a locus of zero axial velocity.

The diameter of the air core varies with the feed volumetric flow rate. If the rate is too low, there is no air core and all of the pulp leaves the cyclone as underflow; if the rate is too high, the air core expands, closing off the apex and forcing all of the pulp to leave the cyclone as overflow. Consequently there is a minimum and maximum volumetric feed rate. Because the pressure drop is proportional to the square of the volumetric feed rate, the minimum and maximum rates can be monitored by the pressure drop. The ratio of the maximum pressure drop to the minimum pressure drop should be less than 4, meaning the maximum to minimum volumetric feed rate should be less than 2.

A particle entering the cyclone finds a point where its velocity with respect to the fluid is equal to the radial velocity, and hence is at rest with respect to radial movement to or from the wall. Because the radial velocity changes with radius, the particle spirals down the cone section, following a path defined

by the particle velocity with respect to the fluid equal to the radial velocity. This path is the equilibrium orbit for all particles of this mass and shape. Larger particles move toward the wall to find their equilibrium orbit, unless they reach the wall. Smaller particles move toward the air core to find their equilibrium orbit. Thus, a particle passes out in the overflow if its equilibrium orbit is to the interior of the locus of zero axial velocity; a particle passes out in the underflow if its equilibrium orbit is toward the outside of the locus. Particles having an equilibrium orbit on the locus define the equilibrium or d_{50} size. As the concentration of particles increases, this simple model (18) breaks down.

For a properly designed and operated cyclone, the sharpness index is constant, typically 0.6. The cut size and apparent bypass are a function of the cyclone geometry, the volumetric feed rate, the material relative density, the feed solids concentration, and the slurry rheology. The relationship for a standard cyclone geometry, where if D_c is the cylinder diameter in cm and inlet area $= 0.05 D_c^2$; vortex finder diameter $= 0.35 D_c$; $0.35 D_c >$ apex diameter $> 0.1 D_c$; and $20^\circ >$ included angle $> 10^\circ$ (19), is

$$d_{50}, \mu\text{m} = 16.9 \frac{D_c^{0.66}}{(\Delta P)^{0.28} (\rho_s - \rho_l)^{0.5} [1 - 1.9 (V)]^{1.43}}$$

(3)

where ΔP is the pressure drop in kPa, ρ_s is the relative density of the solid, ρ_l is the relative density of the liquid, and V is the volume fraction of the solids. This expression has no direct correction for slurry rheology. From this expression, it can be seen that the cut size d_{50} decreases with a decrease in the cylinder diameter and an increase in the pressure drop. Hence, in order to make fine separations, small-diameter cyclones operating at high pressure drops (200–350 kPa (29–51 psi)) should be used. Additionally, the cut size decreases with increasing solid relative density, but increases with increasing solids concentration. One manufacturer, Krebs Engineers (Menlo Park, California) offers cyclones in diameters of 2.5, 10, 15, 25, 38, 51, 66, 76, and 127 cm.

The apparent bypass can be estimated by assuming it is approximately equal to the water split, ie, the percentage of water in the feed that reports to the underflow. The water split has been found to follow a straight-line relationship with the inverse of the feed water rate for cyclones having diameters greater than 7.5 cm and standard geometries. However, for cyclones of smaller diameters, the apparent bypass appears to be much greater than the water split, and is typically proportional to the square root of the water split.

The solids concentration of the underflow must be monitored in order to prevent a roping condition from developing (20). A spiraling solid underflow stream is referred to as a rope discharge. The underflow stream should appear as a hollow cone spray having a 20–30° included angle. Typically, the maximum solids concentration of the feed stream is 30 vol %, the overflow (fine) stream is 15 vol %, and the underflow (coarse stream) is 55 vol %, which would produce a water split of 24%. In order to prevent roping, it is suggested that the underflow stream solids concentration by volume must be less than 49.3 plus one half of the overflow vol % solids.

Because the cut size decreases with cyclone diameter (d_{50} in $\mu\text{m} \sim 5 D_c^{2/3}$ in cm) and because the capacity, Q_s , also decreases with cyclone diameter ($Q_s \sim 0.022 D_c^2$), high classification capacity at low cut sizes can only be achieved by feeding many cyclones of the same diameter. The manifolding of cyclones for parallel processing becomes very important. Each cyclone must be fed at the same pressure using the same amount of solids having the same particle size distribution, therefore radial manifolding is usually recommended in order to achieve uniform distribution of the feed; in-line manifolding usually gives poor distribution.

Centrifuges. Solid-bowl centrifuges have been proposed as an alternative classifying device to hydrocyclones for cut sizes below 10 μm . The results appear to be mixed (21). In one application, where the cut size was 6.5 μm and the sharpness index 0.7, there was essentially no apparent bypass. However, in other applications operating at higher feed concentrations, the cut size ranged from 5–8 μm , but the sharpness index was between 0.3–0.5 and the apparent bypass between 10–30% or higher (22). Smaller cut sizes have also been reported (23).

Dry Classification

Pneumatic classification can be partitioned conveniently into coarse, ie, fine products above 95% < 100 μm ; intermediate, ie, fine products ranging between 95% < 100 and 30 μm ; and fine, ie, fine products below 95% < 30 μm . Pneumatic classification, like hydraulic classification, balances the force of gravity with drag forces (counter flow) in order to bring about a separation.

The vane classifier (Fig. 7) is an example of a free-vortex counterflow classifier. The solids to be classified enter the outer cone dispersed and entrained in a gas stream. A cyclone flow pattern is imparted on the feed stream before it passes through the adjustable vanes into the inner cone. As the vanes are closed down, an increase in the centrifugal motion causes more of the larger particles to strike the inner wall of the inner cone and drop out. The finer particles remain in the gas stream, exiting through the centrally located exit conduit. For a pilot-scale twin-cone classifier, the sharpness index was constant for a set vane position (24). As the vanes were adjusted from 100 (an expansion classifier) to 50 to 25° (a pneumatic cyclone) open, the sharpness index increased but then decreased at the 25° vane setting (Table 1). Increasing the air rate, thereby also increasing the feed rate, caused a decrease in the apparent bypass, but an increase in the cut size for each vane setting.

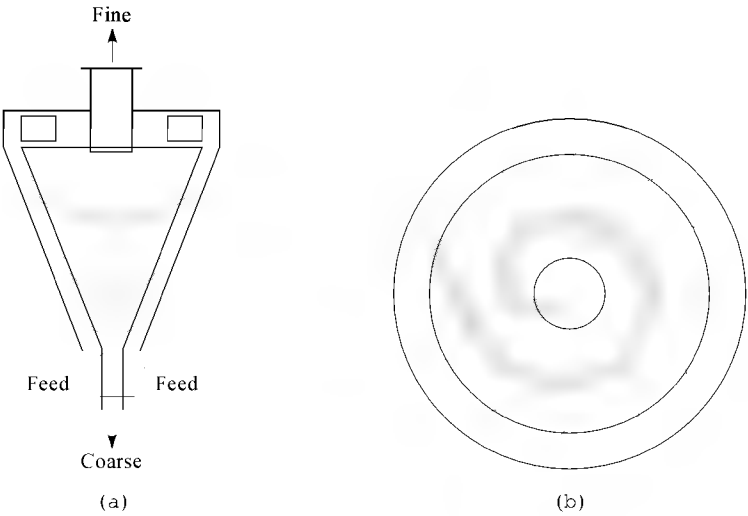


Fig. 7. Vane classifier where (□) represents particle flow: (a) side view and (b) cross-sectional view showing the vanes.

Table 1. Characteristic Classification Parameters for Twin-Cone Classifier^{a, b}

Parameter	Vane setting, °		
	100	50	25

sharpness index	0.325	0.5	0.3
cut size, d_{50} , μm			
1700 m ³ h (1000 ft ³ min)	120	56	22
3400 m ³ h (2000 ft ³ min)	164	105	82
5100 m ³ h (3000 ft ³ min)	197	151	178
apparent bypass			
1700 m ³ h (1000 ft ³ min)	0.05	0.35	0.55
3400 m ³ h (2000 ft ³ min)	0.05	0.20	0.30
5100 m ³ h (3000 ft ³ min)	0.05	0.05	0.05

^a Ref. 24.

^b Diameter = 0.75 m.

The free-vortex principle has been extended to fine classification. A built-in fan in the second chamber of the two-chamber Alpine Mikroplex Spiral classifier draws the classifying air into the classifier chamber while the feed is metered in separately. Adjustable vanes control the angle of air-flow approach to the center of the chamber where the fines flow out. The coarse particles are thrown to the outside and removed mechanically from the casing of the chamber. The sharpness index varies between 0.65 and 0.7, decreasing as the feed rate and hence solids loading increases. The cut sizes range between 10 and 80 μm , depending on the vane setting and the fan speed. No apparent bypass has been reported, but it is appropriate to assume about 5%. This value increases as the feed rate increases.

An alternative design is that of the forced vortex in which the vanes are rotated. The Mikropul Acucut classifier (Fig. 8) draws air in via a vacuum pump. The fines pass through the rotor, discharging out the center. The coarse particles are removed at the outer periphery. The sharpness index varies between 0.6 and 0.8, decreasing as the solids loading increases. The cut sizes range between 2–30 μm , depending on the rpm of the rotor. The d_{50} is inversely proportional to the rpm^{1/4}. Whereas no apparent bypass has been reported, values of approximately 5% have been measured. The forced vortex is less sensitive to solids loading than the free vortex.

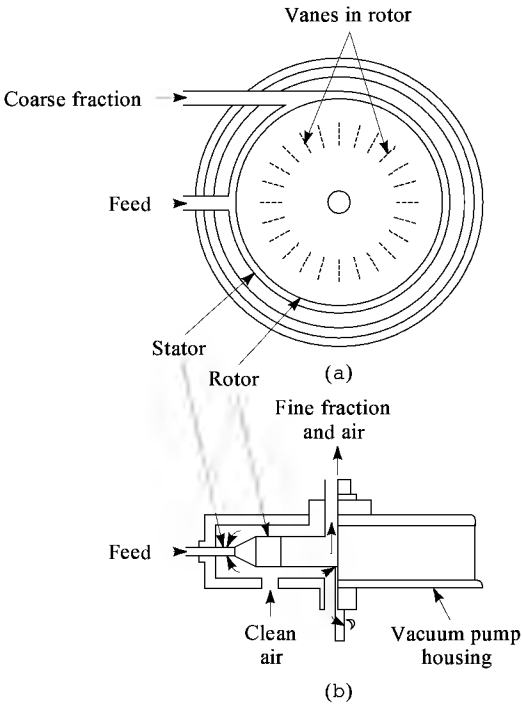


Fig. 8. Views of a Mikropul Acucut classifier: (a) cross section and (b) side view.

The use of a rotor has been the cornerstone of the mechanical air classifier. Rotor design has changed from blades to the multivane or post design (Fig. 9). In this device, the feed material is dispersed in the airstream drawn through the rotor. Whether or not a particle exits in the central fine particle discharge depends on the force balance between the drag force of the particle being conveyed and the centrifugal force created by the rotor against it. The cut size is proportional to the air flow rate raised to a power between 1 and 2 and inversely proportional to the rpm of the rotor and the square root of the relative density of the feed material. Values range between 5 and 150 μm . Operating sharpness indexes can reach 0.75.

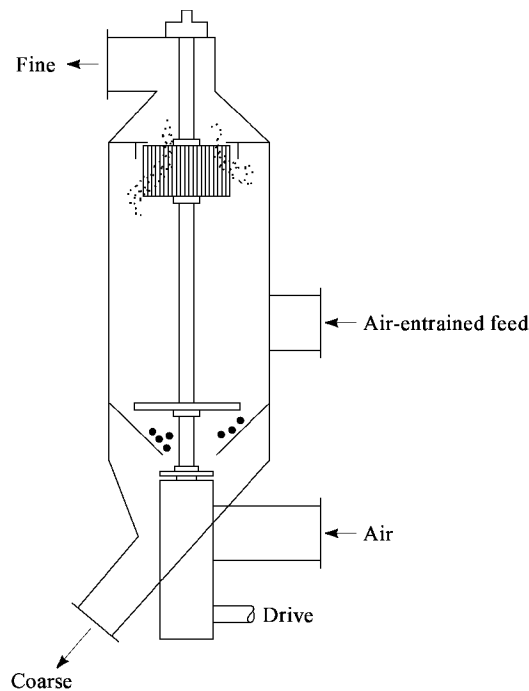


Fig. 9. Tank through-flow classifier.

Data for a pilot device (25) using a blade rotor design are given in Table 2. The sharpness index was a constant 0.6. There are several responses common to classifiers employing rotors. Increasing the feed rate without any other changes reduces the d_{50} value, the efficiency of the separation, and the yield. The data in Table 2 demonstrate a consistent pattern, in all cases, of the d_{50} value decreasing to a minimum with increasing feed rate. Moreover, the efficiency of the classification is reduced with increasing feed rate because the apparent bypass value increases. The yield also decreases with increasing feed rate, because the increased quantity of feed, when going from a relative feed rate of 4 to 7, is merely bypassing to the coarse stream. Assuming the air flow rate to be proportional to fan speed, then the data for a rotor rpm of 1000 give d_{50} proportional to the air-flow rate to the 1.2 power. The 800 and 1400 rotor rpm data show that the d_{50} value is inversely proportional to the rotor rpm. Industrial versions of this separator come in sizes up to 10 m in diameter.

Table 2. Characteristic Classification Parameters for a Mechanical Air Separator^{a, b}

Fan, rpm	Rotor, rpm	Relative feed rate	d_{50} , μm	Apparent bypass	Yield
1400	600	1	62.5	0.05	0.57
1400	600	4	40	0.50	0.20
1400	600	7	40	0.75	0.11
1400	1000	1	32.5	0.10	0.33
1400	1000	4	30	0.75	0.09
1400	1000	7	30	0.80	0.07
1850	800	1	60	0.05	0.55
1850	800	4	40	0.30	0.29
1850	800	7	40	0.40	0.23
1850	1000	1	45	0.05	0.46
1850	1000	4	35	0.30	0.25
1850	1000	7	35	0.50	0.18
1850	1400	1	35	0.05	0.40
1850	1400	4	30	0.40	0.19
1850	1400	7	30	0.60	0.13

^a Ref. 25.

^b Diameter = 1.2 m.

A recirculation design (Fig. 10) returns the gas to the classifier through the fan after the fine particles are removed from the gas stream. Such an arrangement requires an excellent solid–gas separator; otherwise the classification becomes less efficient. A perfect solid–gas separator would be a device having $\alpha = 1$. If the recirculated gas is entered through a secondary coarse stream classification section, then the classification is not less efficient unless the secondary classification is very inefficient.

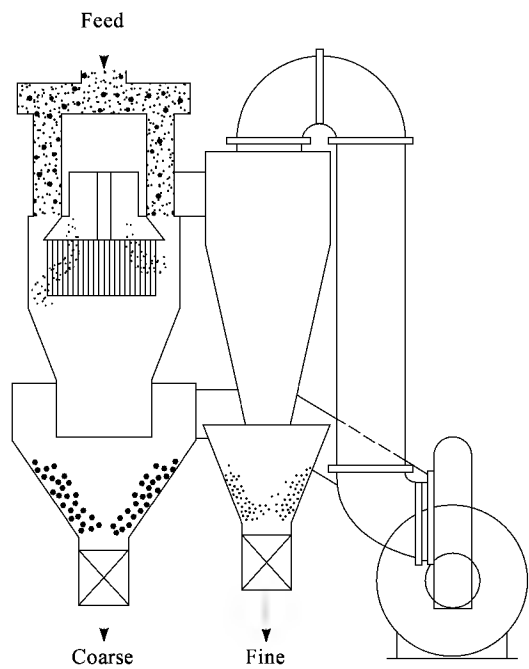


Fig. 10. Recirculating mechanical air separator.

It is quite common in the designs for fine classification to recontact the coarse stream transversely or in counterflow with air before discharging it (see Fig. 9). This removes dry fine particles not removed in the primary classification. That is, these particles are swept back into the feed and given another chance to exit with the fine particles. Such an arrangement increases the overall sharpness index and reduces the overall apparent bypass. Another variation is to reenter the air from the solid–gas separation of the coarse stream.

The Matsuzaka Elbow-Jet classifier (Fig. 11) is based on a transverse flow principle (26). The stream of feed particles are accelerated to minimize the effect of gravity, and introduced into an air jet at right angles. The particles are fanned out in the classification zone with the trajectories for particles of the same hydrodynamic behavior, ie, size and shape, being the same. Classification is achieved by mounting one or more cutters in the classification zone, thus dividing the feed into two or more fractions. A stream of fine particles of less than 5 μm can be produced in this manner.

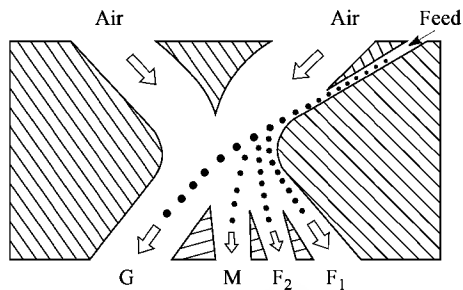


Fig. 11. Cross-sectional schematic of a Matsuzaka Elbow-Jet classifier where F_1 = ultrafine particles, F_2 = fine particles, M = medium particles, and G = coarse particles.

The concept of stage classification in order to improve the classification has been used in both hydraulic and pneumatic classifier installations (27). Different arrangements result in a reduction in the global apparent bypass, an increase in the global sharpness index, or a decrease in the global cut size, compared to the individual units. The underflow–overflow recirculating arrangement (see Fig. 12) can achieve all three global improvements: the sharpness index increases by over 10%; the cut size decreases by 15%; and the apparent bypass decreases by 50%.

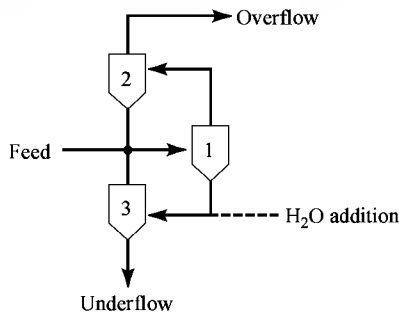


Fig. 12. A triple-stage classifier arrangement.

Health and Safety Factors

The pneumatic classification system should be designed to handle hazardous dust (28). A hazardous dust is one which, when finely divided and suspended in air in the proper concentration, burns, produces violent explosions, or is sufficiently toxic to be injurious to personnel health (see AIR POLLUTION CONTROL METHODS; POWDERS, HANDLING). At the least, almost any dust can be irritating to personnel because of inhalation or skin or eye contact. Fully oxidized and hydrated materials are generally considered safe.

Dust explosions usually occur in pairs. The first explosion involves dust already in suspension. This jars dust from beams, ledges, etc, creating a second cloud to which the explosion propagates, resulting in a secondary explosion. Dust clouds have been ignited by open flames, electric sparks, hot

filaments of light bulbs, friction sparks, hot surfaces, static sparks, spontaneous heating, welding and cutting torches, and other common sources of heat for ignition. Dust-ignition temperatures, for the most part, are 300–600°C. A large majority of the reported minimum spark energies for ignition of dust are <100 mJ (<24 mcal), which is much lower than the energies found in stray electric currents and static electricity discharges. The finer the dust, the greater the concern.

Dust explosions can be minimized by elimination of any leg of the basic fire triangle, ie, source of ignition, oxygen, and fuel. Elimination of the source of ignition and oxygen (<10 wt%) are possible. Dust must be controlled by utilizing air-to-solid ratios throughout the system that keep dust concentrations below a minimum level or above a maximum level. Most organic materials, including plastics, have a minimum explosive concentration of 15–300 g m⁻³, and that of metals can approach 500 g m⁻³.

The equipment in which the dust is handled or stored should be designed to contain the pressure resulting from an internal explosion. Most dusts show maximum pressures of ca 345–700 kPa (50–100 psi); however, the rate of pressure rise changes from ca 700 to 70,000 kPa s (100–10,000 psi s). Equipment-containment design can be coupled with explosive-venting design for the equipment and the building.

Sometimes conveying (qv) velocities, typically 20–25 m s, exceed the flame-propagation rate as determined by tests. Moreover, velocities vary throughout the system. If the flame-propagation rate exceeds the conveying velocities, consideration must be given to the isolation of parts of the system by choking, ie, the use of rotary values, screen conveyors, bins, etc.

Proper ventilation and housekeeping minimizes secondary explosions. Dust collectors of the dry type should be located outside the building, and provided with conduction bags and adequate explosion venting to a safe location.

Prediction of Size Separation

Most corrected characteristic separation curves (27) fit the following logistic function:

$$c(x_i; \kappa, d_{50}) = 1 / (1 + e^{-Y_i})$$
(4)

where Y_i equals $1.0986 (1 - z_i) / (1 + \kappa) - (1 - \kappa)$ or $2.1972 \ln[z_i / \ln \kappa]$ and z_i equals x_i / d_{50} , where x_i is the upper size of the narrow size interval, typically $x_i - x_{i+1} < \sqrt{2}$, hence x_1 is the maximum size. Thus, if the operating sharpness index and cut size are known for a separating device, the size distribution of the fine stream can be predicted. For example, if a feed having the size distribution given in columns 2 and 3 in Table 3 is to be classified, using a device with a sharpness index of 0.6 and a cut size of 100 μm, then using Y_i equal to $2.1972 \ln[z_i / \ln \kappa]$, the corrected values (column 4) can be calculated. There must be a sufficient number of size intervals such that at least one value equals zero and one equals one. Next, the fine stream mass values (column 5) are calculated as the product of the column 3 element times the column 4 element. The fine stream interval values (column 6) are calculated by dividing the column 5 elements by the sum of column 5. Predictions, using a spread sheet, can be used to determine the cut size required to produce the desired product size distribution.

Table 3. Size-Separation Predictions^a

z_i^b	Feed		$c(z_i; \kappa)$	Fine	
	Cumulative	Interval		Mass ^c	Interval ^d
10.0	99.9	0.15	0.00	0.00	0.0
7.1	99.75	1.25	0.00	0.00	0.0
5.1	98.5	4.5	0.00	0.00	0.0
3.55	94.0	11.5	0.00	0.00	0.0
2.5	82.5	16.0	0.02	0.32	0.9
1.8	66.5	16.5	0.07	.15	3.3
1.25	50.0	15.0	0.28	4.20	11.9
0.9	35.0	11.0	0.61	6.71	19.0
0.63	24.0	8.0	0.88	7.04	20.0
0.45	16.0	5.5	0.97	5.34	15.2
0.32	10.5	5.5	0.99	5.45	15.5
0.20	5.0	5.0	1.00	5.00	14.2
Sum				35.21	

^a Using corrected probability of reporting to the fine stream values.

^b $z_i = x_i / d_{50}$. See eq. 4.

^c Product of feed interval and $c(z_i; \kappa)$.

^d Fine mass value divided by summation of masses = 35.21 × 100.

The yield is the sum of column 5 multiplied by (1 - α). Thus, if the operating apparent bypass of the classifier is 0.3, then the yield is 24.65% or 0.2465. The coarse mass values can be calculated by subtracting the fine mass values multiplied by (1 - α) from the feed interval values. The coarse interval values are then calculated by dividing the coarse mass values by the sum of the coarse mass values.

Economic Aspects

The cost of size separation devices varies according to type of separator and the sizing of the components. However, the approximate cost of many of the size-separation devices follow the form,

$$\text{\$} = a X^b$$
(5)

where X is a suitable parameter and a and b are an appropriate coefficient and exponent, respectively. The cost, in U.S. dollars, can be updated by cost index ratio adjustment because the dollars are for a Marshall and Swift mining and milling industry index of 1000 (29).

For vibrating screens, the suitable parameter X is the screen deck area multiplied by the length in m³. Cost includes drive- and feed-boxes, and excludes motor, starter, and screen cloth. For spiral classifiers, the suitable parameter X for costing is the spiral diameter, in cm. Cost includes the motor. For hydrocyclones, the suitable parameter X for costing is the cyclone diameter, D_p in cm. Cost includes fittings and combination urethane–ceramic liners. The appropriate values of X and the coefficients are given in Table 4.

Table 4. Costing Parameters for Size Separators^a

Parameter	X	a	b
-----------	-----	-----	-----

Vibrating screens				
single deck, m ³	0.5–45	9,591.44	0.4069	
double deck, m ³	0.5–45	10,918.50	0.4256	
triple deck, m ³	1.5–45	11,249.90	0.4908	
Spiral classifiers				
simplex, cm	60–125	481.86	1.008	
	125–215	4.24	1.986	
	60–130	611.65	1.043	
duplex, cm	130–200	17.22	1.776	
Wet cyclones				
D _p cm	2.5–35	214.74	0.7582	
D _p cm	35–130	20.01	1.430	

^a See eq. 5.

BIBLIOGRAPHY

"Size Separation" in *ECT* 1st ed., Vol. 12, pp. 520–523, by W. A. Lutz, F. L. Bosqui, and A. D. Camp, The Dorr Co.; in *ECT* 2nd ed., Vol. 18, pp. 366–399, by F. L. Bosqui, Consultant; in *ECT* 3rd ed., Suppl. Vol., pp. 846–871, by P. Luckie, The Pennsylvania State University.

1. T. Allen, *Particle Size Measurement*, 4th ed., Chapman & Hall, New York, 1990.
2. A. M. Gaudin, *Principles of Mineral Dressing*, 1st ed., McGraw-Hill Book Co., Inc., New York, 1967.
3. L. G. Austin, C. H. Lee, F. Concha, and P. T. Luckie, *Min. Met. Process.*, **9**, 161–168 (1992).
4. Equipment Testing Procedure Committee, *Particle Size Classifiers, A Guide To Performance Evaluation*, 2nd ed., AIChE, New York, 1993.
5. R. T. Hukki, in M. J. Jones, ed., *Proceedings of the 10th IMPC*, IMM, London, 1974, p. 240.
6. E. J. Pryor, *Mineral Processing*, Elsevier, New York, 1965.
7. J. P. Nichols, in Mular and Jergenson, eds., *Design and Installation of Comminution Circuits*, AIME, New York, 1982, Chapt. 27.
8. R. J. Batterham, K. R. Weller, T. E. Norgate, and C. J. Birkett, *International Particle Technology Symposium Proceedings*, Amsterdam, the Netherlands, 1980.
9. A. L. Mular, personal communication, 1990.
10. T. Brereton and K. R. Dymott, in Ref. 5, pp. 181–194.
11. J. Slechta, I. J. Taggart, B. A. Firth, and E. Gallagher, *An Evaluation of a Novel Vibrating Fine Screen*, unpublished report, 1981.
12. R. S. C. Rogers, *Powder Technol.* **31**(1), 135 (1982).
13. W. Benson and C. Burgess, *Proceedings 5th ICPC*, Washington, D.C., 1966, pp. 297–311.
14. D. Jenkinson, *Min. Cong. J.* **68**, 29 (July, 1982).
15. E. J. Roberts and E. B. Fitch, *Trans. SME*, 1113 (1956).
16. H. Schubert and T. Neese, in Ref. 5, pp. 213–239.
17. D. F. Kelsall, *Trans. Inst. Chem. Eng. (London)* **30**, 87 (1952).
18. D. Bradley, *The Hydrocyclone*, Pergamon Press, New York, 1965.
19. R. A. Arterburn, in A. L. Mular and G. D. Jergenson, eds., *Design and Installation of Comminution Circuits*, AIME, New York, 1982, Chapt. 32.
20. A. L. Mular and N. A. Jull, in A. L. Mular and R. B. Bhappu, eds., *Mineral Processing Plant Design*, 2nd ed., AIME, New York, 1982, Chapt. 17.
21. P. Scheffler and P. Zahr, *World Mining* **33**(3), 50 (1980).
22. D. Timmermanne and K. Schobert, in J. A. Herbst, ed., *Proceedings 19th IMPC*, San Francisco, 1995, pp. 65–68.
23. H. W. Hennicke and J. Steun, *Mater. Sci. Eng.*, **A109**, 3–7 (1989).
24. Committee on Comminution and Energy Consumption, *National Academy of Sciences Report No. 364*, 1981.
25. L. G. Austin and P. T. Luckie, *ZKG* **29**, 452 (Oct. 1976).
26. H. Rumpf and K. Leschonski, *Chem. Ing. Tech.* **39**, 1231 (1967).
27. P. T. Luckie, in K. V. S. Sastry and M. C. Fuerstenau, eds., *Challenges in Mineral Processing*, SME, Colorado, 1989, Chapt. 13.
28. Technical data, H. Boyd & Associates, Tequesta, Fla., 1981.
29. "Economic Indicators", *Chem. Eng.* monthly.

Peter Luckie
Penn State University

SEPARATION, LOW ENERGY.

See PROCESS ENERGY CONSERVATION; SEPARATIONS PROCESS SYNTHESIS.

SEPARATIONS PROCESS SYNTHESIS

Chemical process and plant design involves a hierarchy of complex and creative activities including both routine and innovative design. Routine design is largely analytical and is primarily concerned with determining values for the specification and operation of specific units such as reactors, extractors, and distillation (qv) columns. By contrast, process synthesis, the generation of conceptual flow sheets comprising such units, is a more open-ended activity characterized by a combinatorially large number of feasible alternatives. Finding better flow sheet alternatives has a significant impact on overall process competitiveness. This is particularly true for separations process synthesis, the selection of separation methods, their interconnection, and their operating parameters. Virtually every chemical process involves the recovery, isolation, or purification of products, by-products, intermediates, wastes, or raw materials. The separation systems to accomplish these tasks often dominate the capital and operating costs of chemical manufacturing processes.

Several approaches to the separations process synthesis problem have been formulated including superstructure optimization, evolutionary modification, and systematic generation. From a known feed composition, desired product compositions, and a well-defined set of separation methods, superstructure optimization approaches construct a hypothetical flow sheet which includes all applicable separation methods interconnected in every possible order so as to include all possible separations scheme alternatives. Separations synthesis then becomes a problem of systematically stripping away the less desirable parts of this superstructure, simultaneously optimizing the design and operating parameters of the remaining separation methods using mixed-integer nonlinear programming (1). This purely mathematical approach is computationally intensive and of limited applicability using the computing hardware and optimization software available as of this writing (ca 1996).

Evolutionary modification starts with an existing flow sheet for a similar separation to which adaptations are made as necessary to meet the objectives of the specific case at hand (2). This approach is exemplified by standard flow sheet patterns, for example, the schemes for breaking heterogeneous minimum-boiling binary azeotropes or the sequences for extractive distillation (see DISTILLATION, AZEOTROPIC AND EXTRACTIVE). Although rarely resulting in novel designs, evolutionary modification is a frequently used separations synthesis technique because of the extensive existing repertoire of design heuristics, standard patterns, and encyclopedias of complete flow sheets.

In the systematic generation approach, the separations flow sheet is constructed from a portfolio of basic components in a directed fashion so that a given feed stream is progressively transformed into one or more target compositions. This process can be viewed as the solution of integrated equipment selection, sequencing, and specification problems. Every separation method is based on differences in one or more physical or chemical properties of the components in a mixture. For a large multicomponent mixture and a moderate number of potential separation methods, the number of possible separations precludes an exhaustive detailed analysis of all options. Resource constraints generally limit the number of separation process alternatives that may be generated and evaluated to a small fraction of the total number of alternatives that are theoretically possible.

As a compromise between thoroughness, efficiency of evaluation, and guarantee of optimality, selection and sequencing methods often are reduced to design heuristics and simple ranked lists of physical and chemical properties characteristic of specific separation methods (3,4). Systematic generation separations synthesis strategies for nearly ideal liquid mixtures as well as quite nonideal liquid mixtures are discussed herein. Strategies involving gas mixtures and dissolved solids are also presented.

Heuristic Distillation Sequencing for Nonazeotropic Mixtures

Problem Representation and Separation Selection. One practical systematic generation problem is the separation of nearly ideal liquid mixtures by simple distillation. By assuming that distillation is to be used for all separations, the separation method selection issue is eliminated. Moreover, if the liquid-phase nonidealities are not severe and the relative volatilities of the components of a liquid mixture are reasonably constant, then knowledge of the boiling points of the components provides sufficient thermodynamic information. The components are ranked in order of increasing boiling point and a split is required when two components adjacent in the ranked list, ie, keys of the separation, are desired to be in different products. Given the additional assumption of sharp separations, ie, high purity and high recovery, the distillate consists of the light key and lower boiling components, whereas the bottoms consist of the heavy key and higher boiling components.

Flow Sheet Construction. Once the components are ranked by boiling point, it is a relatively straightforward task to enumerate exhaustively all possible distillation sequences. Each sequence can then be simulated in detail and the most cost-effective sequence selected by comparison. However, this search approach to separation flow sheet generation makes no use of past process knowledge nor gives any indication of the relative value of any of the sequences without significant computational effort. The use of detailed distillation simulation is a higher level of analysis than is required for screening many of the alternative sequences. Some screening can be accomplished more efficiently with heuristic-based arguments but with the sacrifice of guaranteed optimality.

Heuristics are reliable, well-established rules for reducing the number of potential alternative sequences with minimum effort, and often lead to near-optimal separation system designs. Most of the heuristics for distillation sequencing were originally formulated from parametric studies. A number of Heuristics have been suggested, some of which contradict each other (5–8). Heuristic methods have also been extended to sequencing nonsharp distillation separations and to combinations of distillation, mixing, and stream bypass operations (9–11).

The Sharp Distillation Sequencing heuristics, ordered to help resolve contradictions, are (1) two-product heuristic: if there are only two products in a mixture and all of the components in one product are more volatile than all of the components in the other product, then the next separation should divide the components into two pure products; (2) process hazard heuristic: if a component is corrosive, unstable, reactive, or otherwise hazardous, then the next separation should work toward removing the component from the bulk process stream; (3) direct sequence heuristic: if the most volatile component (product) in the mixture is $>\sim 20$ mol% of the feed, the most volatile component (product) has the largest mole fraction in the feed, the component (product) with the second largest mole fraction is present in a ratio of $<\sim 90\%$ of the most volatile component (product), and the proposed separation is one of the easiest separations remaining, then the next separation should split off the most volatile component (product) as distillate; (4) indirect sequence heuristic: if the least volatile component (product) in the mixture is $>\sim 20$ mol% of the feed, the least volatile component (product) has the largest mole fraction in the feed, the component (product) with the second largest mole fraction is present in a ratio of $<\sim 90\%$ of the least volatile component (product), and the proposed separation is one of the easiest separations remaining, then the next separation should split off the least volatile component (product) as bottoms; (5) distillate-to-bottoms ratio (D/B) heuristic: if a separation has a distillate-to-bottoms ratio (molar basis) of $\sim 40 : 60$ to $60 : 40$ and the proposed separation is one of the easiest separations remaining, then the next separation should be the one closest to a 50:50 distillate-to-bottoms ratio; (6) easiest next heuristic: if none of the other heuristics apply, then the next separation to be done should be the easiest, ie, the one with the highest separation coefficient, S , defined as $S = (\alpha - 1) (D/B)$ for $D/B < 1$, and $S = (\alpha - 1) (B/D)$ for $D/B > 1$, where α is the ratio of volatilities of adjacent

components, D is the molar flow rate of the distillate, and B is the molar flow rate of the bottoms. This series of six heuristics is intended to be examined in sequence. If a heuristic is not applicable, the next one on the list is considered. The difficulty of separation required by some of the heuristics is related to relative volatility and is defined by the following rules for determining difficulty of distillation: if $\alpha > 1.5$ then the given separation is one of the easiest separations remaining; if $1.1 < \alpha < 1.5$ and there are no separations with $\alpha > 1.5$, then the given separation is one of the easiest separations remaining; and if $\alpha < 1.1$ and there are no separations with $\alpha > 1.1$, then the given separation is one of the easiest separations remaining.

The six sequencing heuristics are formulated to reduce the separation load on downstream columns, favoring easier separations early and difficult separations in the absence of nonkey components. If only two products are to be derived from a mixture and all of the components in one product are more volatile than all of the components in the other product, then the next split should divide the mixture into the two products. The presence of hazardous or corrosive materials can greatly increase costs, and such components should be removed as early as possible. The most plentiful product in a mixture should be removed (if it can be) with one separation and if the relative volatility is favorable. Direct sequences, ie, removing a light product as distillate, generally are favored over indirect sequences, ie, removing a heavy product as bottoms. If no product dominates the feed composition, then separations that yield approximately equimolar splits are favored. Only if no other heuristic applies should the easiest separation be performed next.

Example: Fractionation of Fatty Acids. A mixture of waste acids from a vegetable oil refinery are to be separated into five fractions as shown in Table 1.

Table 1. Separation Parameters for Vegetable Oil Acids^a

Product	Component	Concentration, mol %	Bp, at 8.5 kPa ^b , °C	Adjacent relative volatility	Separation coefficient ^c
1	caproic acid [142-62-1]	2	136		
	caprylic acid [124-07-2]	8	165.3	2.45	0.16
2	capric acid [334-48-5]	10	191.3		
	lauric acid [143-07-7]	35	214.6	2.32	1.08
3	myristic acid [544-63-8]	7	236.3	2.23	0.75
4	palmitic acid [57-10-3]	11	257.1	1.68	0.25
5	oleic acid [112-80-1]	12	270.3		
	stearic acid [57-11-7]	15	276.8		

^a Components are listed in order of increasing boiling point.
^b 63.8 mm Hg.
^c Defined by the Sharp Distillation Sequencing heuristics (6).

Four columns are needed to produce the desired products. Considering the Sharp Distillation Sequencing heuristics, heuristic (1) does not apply, as there is more than one product in this mixture. Fatty acids are moderately corrosive, but none is particularly more so than the others, so heuristic (2) does not apply. The most volatile product, the caproic and caprylic mixture, is a small (10 mol %) fraction of the feed, so heuristic (3) does not apply. The least volatile product, the oleic–stearic acids, is 27% of the feed, but is not nearly as large as the capric–lauric acid product, so heuristic (4) does not apply. The split between lauric and myristic acids is closest to equimolar (55:45) and is easy. Therefore, by heuristic (5) it should be performed first. The boiling point list implies that the distillate of the first column contains caproic, caprylic, capric, and lauric acids. This stream requires only one further separation, which by heuristic (1) is between the caproic–caprylic acids and capric–lauric acids.

Returning to the bottoms of the first column containing myristic, palmitic, oleic, and stearic acids, the recomputed mole fractions and volatilities are given in Table 2.

Table 2. Bottoms of First Separation

Product	Component	Concentration, mol %	Adjacent relative volatility
3	myristic acid	16	2.23
4	palmitic acid	24	
	oleic acid	27	1.68
5	stearic acid	33	

The most volatile product (myristic acid) is a small fraction of the feed, whereas the least volatile product (oleic–stearic acids) is most of the feed, and the palmitic–oleic acid split has a good relative volatility. The palmitic–oleic acid split therefore is selected by heuristic (4) for the third column. This would also be the separation suggested by heuristic (5). After splitting myristic and palmitic acid, the final distillation sequence is pictured in Figure 1. Detailed simulations of the separation flow sheet confirm that the capital cost of this design is about 7% less than the straightforward direct sequence.

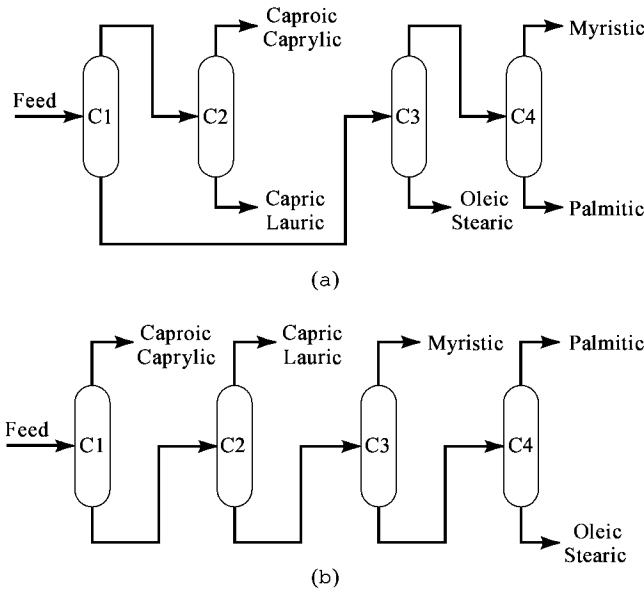


Fig. 1. Fatty acid distillation sequence: (a) sequence generated by ranked heuristics and (b) more expensive direct sequence.

This distillation sequencing problem was solved by an opportunistic systematic generation strategy. At any point in the flow sheet synthesis procedure, the partial design generated is a feasible consequence of the initial feed composition and the separation methods so far specified. The resolution of composition differences between the intermediate streams and remaining product goals are addressed by specifying additional separation methods. No attempt is made to look ahead to anticipate or accommodate any potential difficulties in separation problem resolution. Because distillation is assumed to be feasible for all necessary separations, this opportunistic strategy will not lead to intermediate streams from which it is impossible to resolve the remaining composition differences.

Separations Synthesis for Nonideal Liquid Mixtures

Problem Representation. Typical liquid mixtures encountered in organic chemicals manufacturing often exhibit a wide range of melting and boiling points, reactivity, temperature sensitivity, and strong thermodynamic nonidealities resulting in azeotropism and liquid–liquid phase formation. In general there is a diverse range of behaviors that tend to complicate separation operations. A portfolio of separation methods has been developed over the years to deal with these behaviors, including, for example, simple distillation, azeotropic distillation, dual-feed extractive distillation, decantation, liquid–liquid extraction, and various forms of crystallization (qv), adsorption (qv), and membrane permeation (see [EXTRACTION, LIQUID LIQUID](#); [MEMBRANE TECHNOLOGY](#)). Simple single-feed distillation is most widely used because of predictable, reliable, flexible, robust, and efficient operation and because of mature equilibrium-based design techniques which do not require extensive piloting. Furthermore, simple distillation is one of the few methods which requires only the input of energy to effect the separation. Other common liquid separation methods including extraction, azeotropic distillation, extractive distillation, and solution crystallization require the introduction of an additional mass separating agent (MSA). The MSA must then be recovered and recycled for economical operation, adding further complexity to the separations system design. The generation of separations schemes for liquid mixtures can be thought of as a problem of finding applications for distillation and the identification and resolution of situations where distillation cannot be used.

An important aspect of separations process synthesis is effective problem representation and visualization. The most widely used separation methods for liquid mixtures, including variants of distillation and extraction, are equilibrium-based processes for which the pertinent thermodynamics can be represented conveniently with graphical methods. Triangular phase diagrams (three-component systems) and tetrahedral diagrams (four components) are useful tools for visualizing separation method behavior including material balances. Useful thermodynamic representations are residue curve maps (RCM) and distillation region diagrams (DRD) for vapor–liquid equilibria (VLE), miscibility diagrams for liquid–liquid equilibria (LLE) (see [DISTILLATION, AZEOTROPIC AND EXTRACTIVE](#); [EXTRACTION, LIQUID LIQUID](#)), and solubility diagrams for visualizing solid–liquid equilibria (SLE) (see [CRYSTALLIZATION](#)). Whereas these representations are useful for predicting the behavior of equilibrium-based separations, they do not predict kinetic separation methods such as adsorption and membrane permeation processes, although the material balances for these separation methods still can be represented graphically.

The dominance of distillation-based methods for the separation of liquid mixtures makes a number of points about RCM and DRD significant. Residue curves trace the liquid-phase composition of a simple single-stage batch stillpot as a function of time. Residue curves also approximate the liquid composition profiles in continuous staged or packed distillation columns operating at infinite reflux and reboil ratios, and are also indicative of many aspects of the behavior of continuous columns operating at practical reflux ratios (12).

The family of all residue curves which originate at one composition and terminate at another composition defines a RCM region. All systems with no azeotropes and some systems with azeotropes have only one region, the entire composition space. All residue curves originate at the lowest boiling composition of the system and terminate at the highest boiling composition. However, other systems in which not all residue curves originate or terminate at the same composition have more than one region. The demarcation between regions in which adjacent residue curves originate from different compositions or terminate at different compositions is called a separatrix. Separatrices are related to the existence of azeotropes. In the composition space for a binary system, the separatrix is a point (azeotropic composition); for three components, the separatrix becomes a generally curved line; for four components it is a surface, and so on.

All pure components and azeotropes in a system lie on region boundaries. Within each region, the most volatile composition, ie, either a pure component or a minimum-boiling azeotrope and the origin of all residue curves, is the low boiling or unstable node. The least volatile composition, ie, either a pure component or a maximum-boiling azeotrope and the terminus of all residue curves, is the high boiling or stable node. All other pure components and azeotropes are called intermediate-boiling saddles because no residue curves originate, terminate, or quite pass through these compositions. Adjacent regions may share nodes and saddles. Pure components and azeotropes are labeled as nodes and saddles as a result of the boiling points of all of the components and azeotropes in the system (13). If one component is removed from a mixture, the labeling of all remaining pure components and azeotropes, specifically those which were saddles, may change. Region-defining separatrices always originate or terminate at saddle azeotropes, but never at saddle pure components.

To a first approximation, the composition of the distillate and bottoms of a single-feed continuous distillation column lies on the same residue curve. Therefore, for systems having separatrices and multiple regions, distillation composition profiles are also constrained to lie in specific regions. The precise boundaries of these distillation regions are a function of reflux ratio, but they are closely approximated by the RCM separatrices. If a separatrix exists in a system, a corresponding distillation boundary also exists. Also, mass balance constraints require that the distillate composition, the bottoms composition, and the net feed composition plotted on an RCM for any feasible distillation be collinear and spaced in relation to distillate and bottoms flows according to the well-known lever rule.

For a given multicomponent mixture, a single-feed distillation column can be designed using sufficient stages, reflux, and material balance control to produce a variety of different separations ranging from the direct mode of operation, ie, the pure low boiling node taken as distillate, to the indirect mode, the pure high boiling node taken as bottoms. This range of operability results in a bow-tie shaped set of reachable compositions that may be opportunistically achieved for single-feed distillation roughly bounded by the material balance lines corresponding to the sharpest direct and indirect separation possible. The exact shape of the reachable composition space is further limited by the requirement that the distillate and bottoms lie on the same residue curve, and sometimes further by peculiarities in the shape of the residue curves (14). Because residue curves are deflected by saddles, it is generally not possible to obtain a saddle product (pure component or azeotrope) from a simple single-feed column. For preliminary design purposes, the lowest and highest boiling nodes and the compositions on the distillation region boundary directly opposite the feed composition from these two nodes are of particular interest.

Figure 2 illustrates the situation for the system methyl ethyl ketone (MEK), methyl isopropyl ketone (MIPK), and water, and the problem of recovering a pure MIPK product from such mixtures. The bow-tie approximation of reachable compositions for several feeds is shown in Figure 2a; the exact reachable compositions are shown in Figure 2b.

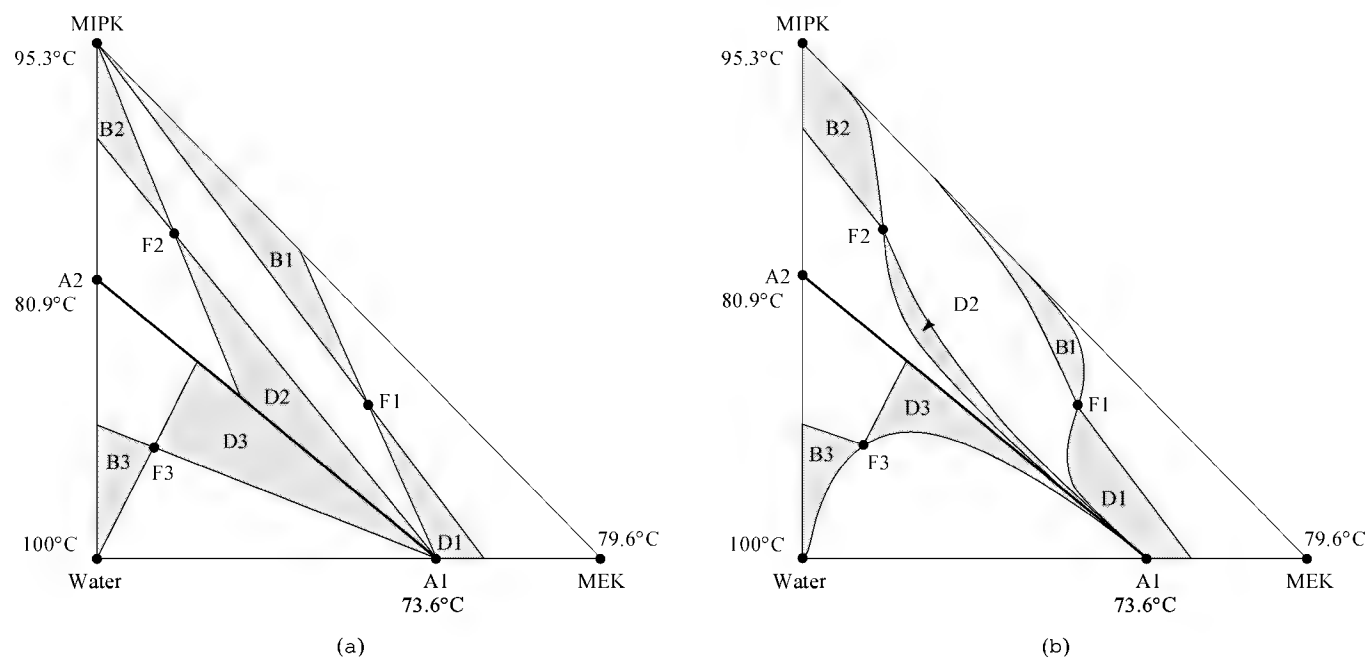


Fig. 2. Methyl ethyl ketone (MEK)–methyl isopropyl ketone (MIPK)–water system where A1 and A2 represent two different azeotropes; F1, F2, and F3, different feed compositions; B*n* and D*n* the corresponding bottoms and distillates, respectively; (—), the distillation boundary; and (■), the reachable compositions for the various feeds: (a) approximate bow-tie and (b) exact reachable compositions.

For feed F1, the upper edge of the reachable composition region is along the MEK–MIPK (water-free) face of the composition triangle and part of the lower edge is along the MEK–water (MIPK-free) face. There exist conditions under which both the water in the bottoms MIPK product can be driven to low levels (high product purity) while the MIPK in the distillate is also driven to low levels (high product recovery), although achieving such operation depends on having adequate stages and reflux ratio.

The reachable composition region for feed F2 is significantly different with the upper edge along the water–MIPK (MEK-free) face of the triangle and the lower edge along the distillation boundary. From this feed it is not possible to achieve a high purity MIPK specification while obtaining high MIPK recovery. If the column is operated to get high purity MIPK, the material balance line is constrained by the distillation boundary. Alternatively, if the column is operated to obtain a high recovery of MIPK, by removing the MEK–water azeotrope as the distillate, the material balance requires the bottoms to lie on the water–MIPK face of the triangle (low purity). From feed F3, which is situated in a different distillation region than the desired product, pure MIPK cannot be obtained by simple single-feed distillation at all.

Phase diagrams are particularly effective tools for separation synthesis work because of their ability to combine various types of thermodynamic information onto one representation including boiling points of pure components and azeotropes, location of azeotropes and any VLE-based distillation boundaries, location of LLE binodal curves and tie lines, melting points, eutectic points, and SLE-phase compositions. Many of these thermodynamic features have a significant impact on the separation system flow sheet. Figure 3 shows examples of combined vapor–liquid and liquid–liquid phase diagrams for several ternary systems. In each of these diagrams a number of special points, lines, and regions have been identified. Figure 3a shows the two-phase liquid region for the MEK–MIPK–water system previously mentioned, whereas Figure 3b illustrates the methylene chloride–2-propanol–water system. Both systems have a distillation boundary spanned by liquid–liquid tie lines. The *n*-hexane–2-propanol–water system is more complicated (Fig. 3c), exhibiting three distillation regions as well as a two-phase liquid region (15).

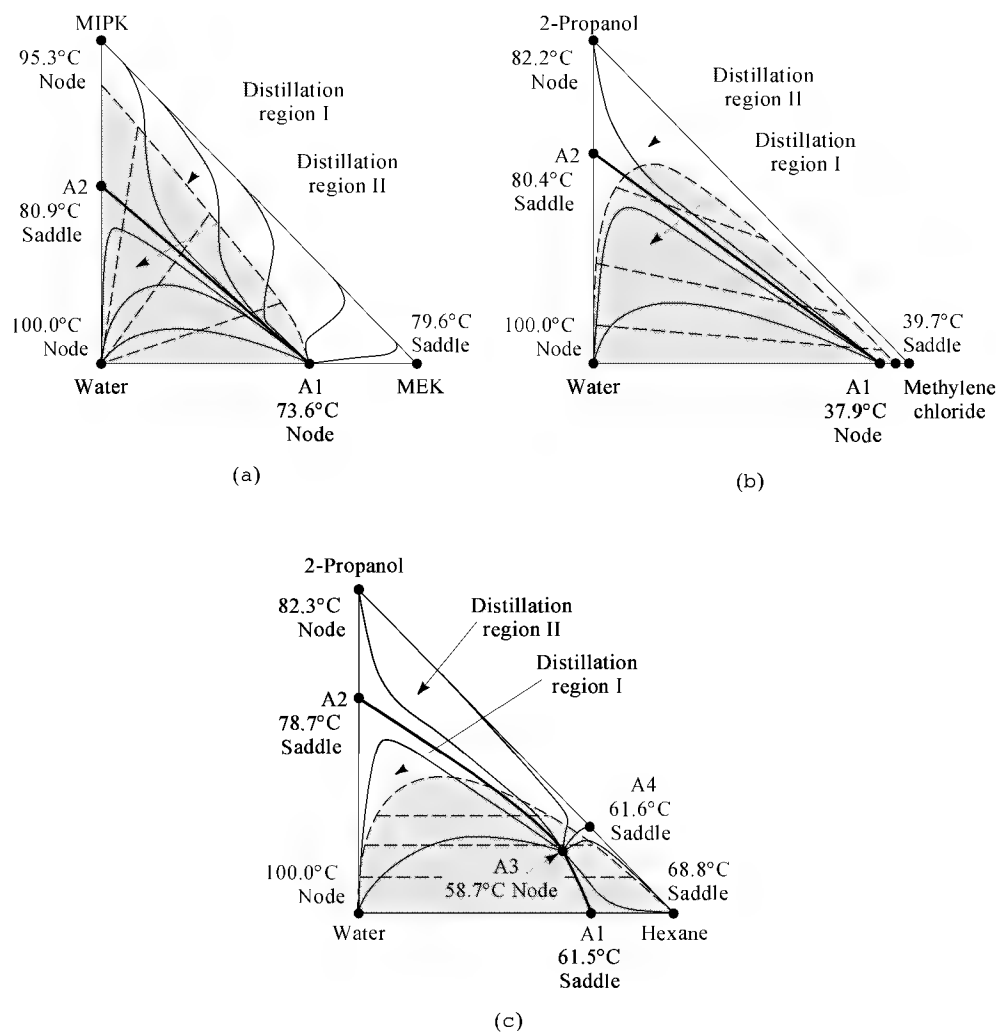


Fig. 3. Combined residue curve and phase equilibria diagrams for the systems where A1–4 represent azeotropes, (—) is the distillation boundary, the shaded areas represent the region of two liquid phases, and () are tie lines: (a) MEK–MIPK–water; (b) methylene chloride–2-propanol–water; and (c) hexane–2-propanol–water.

Relatively few points or composition regions in the phase diagram are of particular significance in separations process synthesis. These compositions of interest include the feed and desired product compositions, as well as azeotropes, eutectics, and selected points on liquid–liquid binodal curves. The choice of composition for a mass separation agent, if required, is critical and usually is a composition which can be conveniently regenerated in the process. Binary and ternary azeotropes which are also high or low boiling nodes, as well as the corresponding two liquid-phase compositions, if heterogeneous, are of particular interest in this regard as these points are reachable compositions by distillation and decantation. Pure components and azeotropes that are saddles are by contrast poor choices for MSA recycle as they are not included in the set of reachable compositions, unless they become nodes when another component in the system is deleted.

Although distillation is a favored separation method, many situations prevent or interfere with the use of simple direct or indirect mode distillation schemes. In general, the designer is faced with avoiding, overcoming, or exploiting a limited set of critical features in order to accomplish the overall objective of the separation system. These critical features include distillation boundaries, ie, if a product (or MSA) composition is in a different distillation region from the feed mixture, the product cannot be obtained directly by simple single-feed distillation; saddle products, ie, if a product (or MSA) is a saddle in a particular distillation region, that product cannot be obtained at high purity directly by simple single-feed distillation; pinched or close-boiling regions, ie, if a feed and product are separated by a region of low relative volatility, simple single-feed distillation is not precluded, but tends to require a large number of stages and high reflux ratio; overlapping melting boiling points, ie, some solutions may contain components with melting points that are higher than the boiling points of other components and distillation of such mixtures may result in solidification within the column; and temperature-sensitive components, ie, some mixtures may contain components that degrade, decompose, polymerize, or otherwise react in an undesirable manner at the temperature conditions of distillation, thus some milder separation method must be used.

Flow Sheet Construction. A purely opportunistic approach to separation synthesis for nonideal mixtures may again be considered. Guided by the general sequencing heuristics given, any separation method may be picked that is feasible and applicable to the state under consideration, and building the solution toward the desired products may proceed. However, the existence of the critical features mentioned previously can preclude the application of simple distillation. In systems exhibiting critical features it is quite possible to reach an intermediate composition from which it is impossible to reduce remaining composition differences by distillation or any other separation method. The only solution is either to backtrack to an earlier intermediate composition, discarding part or all of the given flow sheet, or to follow an alternative synthesis strategy.

The purely opportunistic approach may not make use of all thermodynamic or physical property information known for the whole system. An alternative approach to separation flow sheet generation is to look ahead to potential difficulties and develop contingencies early in the synthesis process to deal with them before running into dead ends. Once a critical feature has been identified, it is useful to examine strategic methods for crossing, breaking, by-passing, reaching, or exploiting the critical feature. The strategies and resulting separation methods for handling a given critical feature may differ considerably and the same strategy can often be implemented in several different fashions. For example, both decantation and extraction are methods for implementing the strategy of exploiting LLE tie-lines to cross a VLE distillation boundary. Table 3 lists strategies and implementations associated with several different critical features.

Table 3. Strategic Separations

Number	Strategy	Separation implementation	Comments
		Pinched regions	
1	use existing liquid–liquid region or add component that causes new or altered	decanter; liquid–liquid extraction	bypasses pinched VLE

	liquid–liquid region		
2	exploit SLE to circumvent cross pinch	melt crystallization; adsorption	^a
3	change system pressure to alter eliminate pinch	distillation	feasibility highly dependent on extent that pinch is altered by moderate change in pressure
4	exploit solid–liquid solubility differences	solution crystallization	^b
5	exploit kinetic phenomena to circumvent cross pinch	molecular sieve adsorption; membrane permeation; pervaporation	requires the selection of solid-phase mass separation agent
6	use VLE to circumvent cross pinch; approach pinched region from non-pinched direction	distillation; azeotropic distillation	choose third component which does not form pinch with one of the key components
7	distill through pinched region	simple distillation	may require many stages and high reflux ratio
8	use existing component or add new MSA to bypass pinch in two-feed column	extractive distillation	^c
		Saddle products	
1	^d	^d	if binodal composition not pure enough for final product, additional purification may be difficult
2	exploit SLE to reach saddle	^e	^e
3	use existing component or add new MSA to reach saddle in single two-feed column	^f	^f
4	exploit solid–liquid solubility differences to reach saddle	^g	^g
5	reduce expand problem dimensionality to turn saddle into a node; move to face of RCM	azeotropic distillation; distillation	three-component system saddle becomes two-component system node
6	exploit kinetic phenomena to reach saddle	^h	^h
7	react to turn saddle into node	reaction; reactive distillation	reactions involve existing components
		Overcoming distillation boundaries	
1	^d	^d	uses LLE to avoid boundary in VLE
2	exploit SLE	^e	^e
3	mix streams to cross boundary	mixer	may generate infeasible solutions because of mass balance violations
4	exploit existing or add component that causes favorable convex curvature to cross boundary	distillation	curvature difficult to predict accurately without experimental data; must be highly curved to avoid large recycles
5	change system pressure alter eliminate boundary	ⁱ	ⁱ
6	^g		
7	exploit kinetic phenomena to cross boundary	^h	^h
		Vapor–solid–liquid equilibria	
1	use intermediate reflux of low mp component to prevent freezing of high meter	sidedraw-reflux column	occurs when bp and mp of components overlap
2	exploit SLE to separate overlapping melters boilers	melt crystallization	^a
3	add component to remove lighter components to separate overlapping melters boilers	stripping	
4	exploit solid–liquid solubility differences to separate overlapping melters boilers	^g	^g
		Temperature-sensitive mixtures	
1	^d	^d	perform separation at conditions below reaction–degradation temp (20–45°C)
2	exploit SLE	^e	^e
3	alter system pressure to decrease reboiler temp	vacuum distillation; wiped-film and short-path evaporation	feasibility highly dependent on pressure required to prevent degradation and stage requirements of desired separation
4	^g	^g	^b
5	exploit kinetic phenomena	molecular sieve adsorption; membrane permeation	^h

^a Driving force for melt crystallization is melting point (mp) difference of impurities and desired product; feed must meet mp and crystal characteristics for feasibility (see also solution crystallization).

^b Based on differences in melting points and liquid-phase solubilities; four modes of operation possible: drown-out, isothermal evaporation, adiabatic evaporation, and cooling (choice depends on stream characteristics).

^c Extractive agent modifies liquid-phase behavior (activity coefficients) of key components; RCM must be of appropriate form for extractive distillation to work.

^d Same as Strategy 1 of *Pinched regions*.

^e Same as Strategy 2 of *Pinched regions*.

^f Same as Strategy 8 of *Pinched regions*.

^g Same as Strategy 4 in *Pinched regions*.

^h Same as Strategy 5 of *Pinched regions*.

ⁱ Same as Strategy 3 of *Pinched regions*.

Early in the synthesis of separation schemes for nonideal liquid mixtures, it may not be known exactly where in the flow sheet a strategic separation is going to be, only that it is required some place in some form in order to overcome or exploit a particular critical feature. Thus, strategic separations often

initially do not have well-defined feeds and products. The region where the feed must be located may be known, as well as a general idea of the types of products expected, but no definite compositions for either. For example, for a strategic decant operation, all that may be known initially is that the feed must be somewhere in the two-phase liquid region and the products are on the binodal curve at opposite ends of a tie-line through the feed possibly specified to extend into two different distillation regions.

Along with the forward-looking strategic approach, several opportunistic separations may also be possible. These include (1) distill the low boiling node of a mixture overhead, ie, simple direct distillation sequence (bottoms composition is constrained by mass balance and region boundary); (2) distill the high boiling node of a mixture as underflow, ie, simple indirect distillation sequence (distillate composition is constrained by mass balance and region boundary); and (3) decant a two-phase liquid mixture. When critical features are present, opportunistic operations can be thought of as links between feeds, strategic separations, and products. An opportunistic operation often is used to reach a composition where a strategic separation is applicable, eg, opportunistically distill into a two-phase liquid region, then strategically decant the mixture to cross a VLE boundary. Sometimes an opportunistic separation is equivalent to a strategic separation. Alternatively, if no critical features are present, the entire separations flow sheet can be synthesized heuristically by a series of opportunistic separations alone.

Separation Synthesis Algorithm. The procedure for the systematic generation of conceptual separation system flow sheets consists of eight steps. The synthesis scheme emphasizes the use of the appropriate knowledge, identification of critical features and strategies to deal with them, and provides a means for selecting and sequencing both opportunistic and strategic separations. The process begins by the construction of lists of sources and destinations for the process. A source may be the original feed mixture or a stream created by a strategic or opportunistic separation. A source list is maintained as the algorithm progresses. It contains streams which have not been identified as destinations, have not been recycled, or fed into another unit operation. Destinations are the final and sometimes intermediate objectives of the separation flow sheet. They may be products, by-products, MSA compositions that must be regenerated and recycled, or the feed to a strategic separation. The destination list also changes as the design proceeds.

The separations synthesis algorithm is as follows.

Step 1. Define problem.

- Determine feed specifications: construct list of all sources for process (source list initially consists of all feeds).
- Determine product specifications: construct list of all destinations of process components (destination list initially consists of compositions and amounts of all desired products and by-products to be produced by the separations process).
- Determine identity of any additional species to be used as mass separating agent (MSA).
 - Decision points — product specifications and identity of MSA species

Step 2. Examine equilibrium data.

- Examine VLE, LLE, and SLE as needed.
- Determine how each varies with temperature, pressure, composition, etc.
- Construct a residue curve map (RCM) with LLE and SLE overlaid.
- List compositions of interest, ie, potential MSA compositions.
- Select composition specification for MSA (if any) and add composition to destination list.
 - Decision point — MSA composition

Step 3. Choose stream to process.

- Choose stream from source list to process (if list contains >1 stream, selection may be arbitrary); all streams will eventually be processed, but processing order may have effect on structure of flow sheet synthesized.
- Prepare list of known critical features which must be dealt with to reach all destinations (products as well as MSA compositions).
 - Decision point — selection of current stream to process

Step 4. Process stream.

- Consult Recycle heuristics and Rules for Selecting Operations (tiers 1 and 2) and do one of the following with the current stream:
 - (1) If composition matches that of a destination, label stream as product (note that destination has been reached at least once).
 - (2) Recycle stream (consult Recycle heuristics for guidance).
 - (3) Specify operation from list of applicable opportunistic and strategic operations.
 - Determine if any of the previously identified critical features pertain to current stream.
 - Examine Table 3 (corresponding to critical features identified) and list all strategic operations applicable to given stream.
 - List all opportunistic separations applicable to given stream (see Table 3).
 - Determine which strategic or opportunistic operation should be tried first.

Decision points = selection of destination, recycle point, and operation to be applied to given stream

Step 5. Update flow sheet.

- Connect streams.
- Update flow sheet.
- Add any newly created streams to source list, ie, streams resulting from separations just added to flow sheet.
- Add required feed stream resulting from strategic operations just specified to destination list.
- Update worksheet (stream, possible moves, chosen move, decision point).
 - Decision points = none

Step 6. Other streams to process?

- If source list is not empty, then flow sheet is not complete and additional streams must be processed.
 - Return to *Step 3*.

Step 7. All product (including MSA and strategic operation feed) specifications met?

- Solution is complete if all product specifications (specs) met.
- If flow sheet cannot meet the product specs as originally stated, determine if they can be relaxed. Are products specs which can be obtained with given flow sheet actually satisfactory? If yes, solution is complete.
 - Decision point — all problem specs met?

Step 8. Generate alternative solutions.

- Once acceptable flow sheet is generated, consider evolutionary modification.
- If another solution is desired, return to any all decision points and choose different alternatives.

The thermodynamics and physical properties of the mixture to be separated are examined. VLE nodes and saddles, LLE binodal curves, etc, are labeled. Critical features and compositions of interest are identified. A stream is selected from the source list. This stream is either identified as meeting all the composition objectives of a destination, or else as in need of further processing. Once an opportunistic or strategic operation is selected and incorporated into the flow sheet, any new sources or destinations are added to the respective lists. If a strategic separation for dealing with a particular critical feature has been implemented, then that critical feature is no longer of concern. Alternatively, additional critical features may arise through the addition of new components such as a MSA. The process is repeated until the source list is empty and all destination specifications have been satisfied.

Separation method selection is governed by a two-tier set of rules for selecting among potential operations. The first tier involves the selection between strategic and opportunistic operations. Strategic separations are favored, as these are known to be required in the flow sheet at some point, as are separations which directly reach a desired product composition. The rules in the first tier state that once all the potential opportunistic and strategic operations have been identified for the current stream, the next step is to determine the order in which these operations should be tried. The following

heuristics may be used to order alternatives at a decision point: (1) select a strategic separation which reaches a product (or MSA) composition; (2) select an opportunistic separation that reaches a product (or MSA) composition; (3) select a strategic operation that does not reach a product; (4) select an opportunistic operation that reaches feed composition of a strategic operation; and (5) select any remaining opportunistic separation. The second tier uses the same general sequencing heuristics outlined previously to help guide the process toward lower cost sequences when more than one alternative is feasible. These are (1) perform hardest separations last; (2) favor separations which remove corrosive, hazardous components as a product directly, ie, no further separation required for that stream; (3) favor separations which remove the largest fraction of mixture as a product directly; and (4) favor separations which give approximately equal-sized output streams.

The recycling of a stream to a point upstream in the process is often a powerful alternative to performing additional processing. In particular, regeneration and recycle of an MSA composition is essential to the economic operation of such separation methods as extraction, extractive distillation, and azeotropic distillation. However, recycling cannot be done indiscriminately and material balances must be carefully considered. Unit operations which worked before closing the recycle may no longer function in the same manner, or may become infeasible. For example, the recycle may cause an upstream composition to move into another distillation region making a previously specified distillation impossible. Alternatively, the recycle may result in the infinite buildup of a particular component if there is no outlet for that component from the recycle loop. The Recycle heuristics are (1) consider recycling a stream when the separation(s) to be performed on that stream would duplicate a section of the process that has been previously specified, which generally occurs when the composition of a stream is similar to the composition of a previous stream in the process; (2) avoid recycle to a point that would allow infinite buildup of any component, ie, if there is no outlet for each component downstream of a recycle, then do not recycle to that point; (3) similarly if a stream is the last stream on the source list and all products have not been obtained, recycle is not an option; (4) match stream compositions as much as possible; (5) avoid recycle of streams that cause significant changes in the operation of upstream units, eg, do not recycle a stream at a point that would cause an upstream composition to move into another distillation region; (6) if compositions do not match, and heuristic (5) is not violated, then it is permissible to recycle a stream which is small compared to an upstream flow rate; (7) if all product specs have been met and all critical features dealt with, and there are still streams on the source list, then consider recycling remaining streams; and (8) do not add a new MSA if all products and MSA compositions of interest have been reached and there are still streams on the source list (consider recycling remaining streams).

Separation Selection Issues. Separations process synthesis is an open-ended design activity. Often very distinct flow sheets can be generated if the choice of MSA is changed, a different separation method is applied to a particular stream, or even by changing the order in which streams on the source list are examined. Such decision points are found throughout the algorithm and are the mechanism for generating alternative flow sheets. Each decision point can be revisited, another alternative selected, and the section of the flow sheet resulting from this decision point can be revised.

It is particularly difficult to give rules for finding the best separation method among several feasible alternatives for nonideal mixtures. Ranked lists of physical properties may not have enough discriminating power without resorting to more in-depth calculations. However, knowledge of a few properties characteristic of each separation often can help eliminate feasible but poor choices as well as indicate clearly better alternatives. Sources of information on MSA selection are included in References 2 and 16–19. Table 4 lists pure component and mixture properties which characterize several liquid separation methods. General considerations for the various distillation-based separation methods, crystallization methods (20–22), and other methods requiring MSA (16,23–25) are as follows.

Distillation

- (1) Usually cannot be beaten when α (relative volatility) >1.5 , competitive for $1.1 < \alpha < 1.5$, but other methods usually better for $\alpha < 1.05$ – 1.1 .
- (2) Becomes difficult (may require many stages or reflux) when temperatures of high and low boiling nodes of region differ by $<10^{\circ}\text{C}$.
- (3) Very difficult when temperatures of high and low boiling nodes of region differ by $<5^{\circ}\text{C}$.
- (4) Simple one-column distillation is generally unfavorable when the distillate-to-bottoms ratio (D/B) is either very high or very low, eg, $95 : 5 < \text{D/B ratio} < 5 : 95$. For low D/B ratio high reflux required; for high D/B ratio large amount of feed must be vaporized.
- (5) Consider special multiple-effect distillation schemes if the D/B ratio is either very high or very low, or if the relative volatility is low.
- (6) Avoid attempts to recover simultaneously both high and low boiling nodes in high purity from mixtures of >3 components, particularly in columns that reflux compositions different from the distillate composition, ie, reflux of one phase from a decanter, as such operations may be difficult to control.

Azeotropic distillation

- (1) The more structurally, chemically similar components are, the less likely that the separation will be improved by azeotropic distillation (if an MSA-key component azeotrope is being used to alter the RCM); any azeotropes formed between one component and another similar component tend to have similar boiling points, compositions.
- (2) MSA selection is critical.
- (3) See step 6 for *Distillation*.

Extractive distillation

- (1) Usually not favorable for separation of components that show similar liquid-phase behavior, ie, stereoisomers, homologous series, *iso*-normal-*neo* isomers; components to be separated must have some different functional group for MSA to affect liquid-phase behavior differently.
- (2) Only certain RCMs are favorable for extractive distillation.
- (3) MSA selection is critical.

Pressure-swing distillation

- (1) Azeotrope composition must change at least 5–10 mol % over moderate pressure change for process to be economical. Generally larger change required for minimum than maximum boiling azeotrope. Extent of pressure change somewhat dictated by reboiler temperature in high pressure column, condensing temperature in low pressure column (avoid refrigeration, special heat-exchange fluids).
- (2) Generally less favorable for minimum boiling azeotrope, because the recycles between columns are taken overhead, pure products are bottoms; results in high energy usage and larger column size; products contain any high boiling contaminants.
- (3) Generally more favorable for maximum boiling azeotrope because the recycles between columns are bottoms streams, pure products are distillates; recycle not as energy-intensive, products distilled once.

Distillation into curved boundary

- (1) More curved the boundary, less recycle required (generally less favorable for minimum boiling azeotropes because of large overhead recycles).
- (2) Consider other methods if recycle rates are high, ie, >3 – 5 times feed rate.
- (3) Boundaries involving maximum azeotropes often highly curved.
- (4) Not applicable to binary systems; azeotrope is point in phase diagram.

High vacuum distillation

- (1) Consider high vacuum distillation when material is temperature-sensitive at normal pressure ranges.
- (2) Consider wiped-film evaporator (WFE) (external condenser) when pressure must be in 0.6–0.013 kPa (5–0.1 torr) range.
- (3) Consider short-path evaporator (SPE) (internal condenser) when pressure must be in 0.013–0.00013 kPa (0.1–0.001 torr) range; for larger-scale production, 0.013–0.0007 kPa (0.01 to 0.005 torr) is practical lower pressure limit.

- (4) At best only one equilibrium stage achievable for each WFE or SPE unit; separation efficiency decreases as pressure decreases.
- (5) Operate at a vacuum no lower than required; SPE unit typically 25–35% higher cost than WFE unit.
- (6) Separate all noncondensables before feeding to WFE or SPE unit (even a small amount of noncondensables overloads vacuum system, especially at ultrahigh vacuum ranges); most low molecular weight compounds do not condense at cooling water temperatures under high vacuum.

Solution crystallization (general)

- (1) Cost increases greatly with decreasing temperature (keep ≥ 0 –20°C if possible).
- (2) Order of preference: adiabatic evaporation $\succ$ isothermal evaporation $\succ$ cooling $\succ$ down-out.
- (3) If system forms solid solution, multistage crystallization required (likely to be expensive).

Solution crystallization (adiabatic evaporation (vacuum cooling))

- (1) Favored for systems where temperature change has moderate effect on solubility.
- (2) Generally best when crystallizing component is 20–75% of feed mixture.
- (3) Solvent must have sufficient volatility to be easily removed at target temperature and vacuum levels.
- (4) May be inferior for systems with high boiling point elevations (high vacuum levels required).
- (5) Components must be stable at boiling point of mixture (probably run under vacuum); vacuum lower than 7–13 kPa (50–100 torr) may require specialized (scraped-film) crystallizer, more complicated vacuum system.
- (6) If running under vacuum, separate noncondensables before feeding to crystallizer (excessive noncondensables overload vacuum system or lead to oversized design).
- (7) Complete recovery possible in theory (in practice, recovery dictated by solubility limit of impurities and viscosity of solution).

Solution crystallization (isothermal evaporation)

- (1) Favored for systems where solubility has little temperature dependence, but can be used in other cases.
- (2) Solvent to be evaporated must have high volatility compared to remainder of solution.
- (3) Theoretical yield is 100%, but usually limited by cocrystallization of impurities.
- (4) Use single effect if crystallizing component is $\geq 75\%$ of feed.
- (5) Consider multiple effects if crystallizing component is $< 75\%$ of feed.
- (6) Energy usage per kg of product highly dependent on heat of vaporization of solvent and number of effects used (can be very energy intensive).
- (7) See considerations (5)–(7) of *Vacuum cooling*.

Solution crystallization (cooling)

- (1) Generally best when crystallizing component is 20–75% of feed mixture.
- (2) Solubility should be strong function of temperature (should decrease with decreasing temperature).
- (3) Feed should be close to saturation limit before cooling to maximize potential recovery (consider preconcentration step to remove excess solvent).
- (4) Recovery limited by eutectics.

Solution crystallization (down-out/salt-out)

- (1) Consider when solvent cannot be removed by evaporation (either temperature sensitivity or low solvent volatility).
- (2) For down-out, look for MSA in which the component to be crystallized has low solubility, while other components have high solubility (MSA should be completely miscible with feed mixture).
- (3) For salt-out, add solute which reduces effective solubility of component to be crystallized (salt having common ion is usual choice); salting-out compound should be cheap, as may be difficult to recover.
- (4) Generally best when crystallizing component is large percentage of feed (consider preconcentration step if dilute).

Melt crystallization

- (1) Mixture must be stable at temperatures 20°C above melting point of crystallizing component.
- (2) Crystallizing component should be at least 50% of feed mixture, preferably $\geq 75\%$.
- (3) Melting points must be between 30 and 150°C for practical implementation.
- (4) Crystallizing component should show sharp melting point.
- (5) Metastable zone width should be less than 25°C for efficient crystallization.
- (6) Large difference in melting point and eutectic temperature improves controllability.
- (7) Difference of 20–30°C between melting points of pure components more favorable.
- (8) Recovery limited by eutectics.
- (9) Cost increases greatly with decreasing temperature (keep ≥ 0 –20°C).

Stripping (steam)

- (1) Particularly good for components that form minimum boiling azeotropes with water and are immiscible with water; tend to have extremely high volatilities.
- (2) Not good for low boilers that are completely miscible with water, eg, methanol or acetone.
- (3) Generally uneconomical if component to be stripped is more than 10–15% of feed mixture (too much steam required); should use other methods to remove bulk of component.
- (4) Expect some product contamination if feed components can react with water, eg, ester will be partially hydrolyzed to acid and alcohol; fate of reaction product species depends on above rules, eg, methanol from methyl ester hydrolysis probably not stripped out of bottoms stream.

Extraction

- (1) Usually not favorable for separation of components which show similar liquid-phase behavior.
- (2) MSA selection is critical.
- (3) Consider component already in mixture for MSA if system exhibits appropriate liquid–liquid behavior.
- (4) For given binary pair, two-phase behavior likely when infinite dilution activity coefficient of either component in other component is ≥ 7.5 .
- (5) For trace removals, eg, catalyst recovery, usually to < 100 ppm; often difficult to regenerate solvent to such low levels.

Adsorption (liquid-phase)

- (1) Species separated by molecular sieving effects when kinetic diameters fall into different zeolite aperture size categories (standard molecular sieve diameters 300, 400, 500, 800, 1000, 1300 pm).
- (2) Dehydration of organics (removal of $< 1\%$ water) generally feasible by molecular sieving, if kinetic diameter of organic ≥ 300 pm.
- (3) Selectivity generally unfavorable when system components miscible in all proportions (may not be problem when removal is objective).

- (4) Sparingly soluble minor components are more favorably adsorbed.
- (5) Strong hydrogen bonding among components retards adsorption.
- (6) Generally for dilute solutions (<1–2% of feed), unless molecular sieving effect also present.
- (7) Activated carbon adsorption generally uneconomical for removal of >1000 ppm contaminant from large stream unless bed regenerated; steaming often easiest regeneration method but creates new wastewater problem; usually 3–5 kg steam required per kg of carbon for regeneration.
- (8) Do not regenerate molecular sieves by steaming; water typically is strongly adsorbed and may not be easily displaced by adsorbent in next adsorption cycle.
- (9) Resin adsorbents (macroreticular polymer resins) generally good for removal of up to 1–2% of stream (often regenerable).

Membrane permeation and pervaporation

- (1) Membranes becoming more widely available for aqueous–organic separations; some successful industrial applications reported for dehydrations and removal of alcohols (ethanol and above) from water.
- (2) Not good for services where solids, tars are expected; stream should be distilled once before coming in contact with membrane.

Table 4. Characteristic Properties for Liquid Separation Methods^a

Separation method	Characteristic properties	Separation principle	Separation agent	Industrial application
Distillation-based methods				
distillation	relative volatility; D/B ratio of mixture; RCM or DRD	difference in volatility	heat transfer	separation and purification of gasoline, kerosene, fuel oils, other petroleum (qv) products
extractive distillation	data, model for liquid-phase behavior; RCM or DRD	difference in volatility as altered by MSA	heat transfer and solvent	separation of toluene from close-boiling nonaromatic hydrocarbons ^b
azeotropic distillation	RCM or DRD	difference in volatility and related azeotrope formation	heat transfer, possibly added entrainer	dehydration of ethanol (qv)
pressure-swing distillation	effect of pressure on azeotrope composition; type of azeotrope ^c	difference in volatility; effect of pressure on azeotrope formation	heat transfer	dehydration of tetrahydrofuran
distillation in-to curved boundaries	detailed, accurate RCM	difference in volatility; related azeotrope formation	heat transfer, possibly added entrainer	production of strong nitric acid (qv)
high vacuum distillation	effect of pressure on bp; relative volatility	difference in volatility	heat transfer	production of polycarboxylic acids from long-chain un-saturated fatty acids
Crystallization methods				
adiabatic evaporation	effect of temperature on solubility; bp and heat of vaporization of solvent	difference in freezing tendency, solubility	heat transfer	production of urea (qv)
isothermal evaporation	effect of temperature on solubility; bp and heat of vaporization of solvent	difference in freezing tendency, solubility	heat transfer	production of refined sugar from sugar beets (qv)
cooling	effect of temperature on solubility; eutectic diagram	difference in freezing tendency, solubility	heat transfer	manufacture of potassium nitrate
drown-out or salt-out	solubility in solvent	difference in freezing tendency; solubility in added solvent; salt solution	heat transfer, added solvent or salt	purification of temperature-sensitive pharmaceuticals (qv)
melt crystallization	mp; eutectic diagram; metastable zone width	difference in freezing tendency	heat transfer	separation of <i>meta</i> - and <i>para</i> -xylenes
Other MSA methods				
steam stripping	miscibility, azeotrope formation, and reactivity with water	difference in volatility	heat transfer	removal of volatile organics from wastewater
extraction	data, model for liquid-phase behavior; selectivity in solvent	difference in partitioning behavior between two liquid phases	liquid solvent	separation of toluene from nonaromatic hydrocarbons
adsorption	kinetic diameter; hydrogen bonding characteristics	difference in affinity for adsorption on solid surface	solid adsorbent, possibly aided by temperature or pressure gradient	separation of <i>meta</i> - and <i>para</i> -xylenes
pervaporation	chemical families	forced flow through semipermeable membrane	membrane	removal of low levels of water from organics

^a D/B = distillate-to-bottoms ratio, RCM = residue curve map, DRD = distillation region diagram, and MSA = mass separating agent.

^b See BTX PROCESSING.

^c Maximum or minimum boiling.

Solution Evolution. It is often beneficial to re-examine a completed flow sheet and look for opportunities for simplification and consolidation of unit operations. A complicated series of unit operations can sometimes be replaced by a simpler structure that has equivalent material balance lines. When two or more sections of the flow sheet perform similar functions, ie, both produce the same product using the same or similar unit operations, one section often can be eliminated by recycling the stream to the input of the remaining section. An MSA contaminated by other components in the mixture often functions as effectively as a pure MSA without the need for additional purification operations.

Example: Separation of 2-Propanol–Water. Consider the separation of a binary mixture of 60 mol % water and 40 mol % 2-propanol (or isopropyl alcohol (IPA)) into two products consisting of 2-propanol of 99.5 mol % purity and water with <100 ppm 2-propanol impurity.

Binary System. The first task is to examine the characteristics of the 2-propanol–water-phase equilibria (VLE, LLE, SLE) to determine the compositions of interest and any critical features. 2-Propanol forms a minimum boiling azeotrope with water (80.4°C at 101.3 kPa (760 torr), 68 mol % 2-propanol). The azeotrope is between the feed and the IPA product and is a distillation boundary, thus it is impossible to obtain both desired products from any single-feed

distillation column.

Table 3 contains strategic separations to be considered for crossing distillation boundaries. Many of these can be eliminated after examining the pertinent physical properties and equilibrium behavior (see Table 4) and referring to the general separation considerations. The results are summarized in Table 5.

Table 5. Strategic Separations for 2-Propanol–Water System

Strategy ^a	Status	Comments ^{b,c}
1. use existing liquid–liquid region or add component that causes new or altered liquid–liquid region	extraction, decantation infeasible with current system	water and IPA miscible in all proportions (no liquid–liquid region in binary system)
	may be feasible with addition of suitable MSA	components show different liquid-phase behavior (exploitation of LLE pos-sible with appropriate MSA)
2. exploit solid–liquid equilibrium	melt crystallization probably inferior	possibly feasible, but freezing points of water and IPA (273 and 184.7 K, respec-tively) low
	adsorption infeasible	strongly hydrogen bonded, miscible system (adsorption unfavorable)
3. mix streams to cross boundary	mixing infeasible	binary azeotropes cannot be broken in practice by simple mixing (requires infinite recycle or mixing of IPA purer than desired product)
4. exploit existing curva-ture or add component to cause favorable curvature to cross boundary	infeasible alternative; may be feasible by adding suitable third component	no convex curvature to exploit because azeotrope is point in phase diagram
5. change system pres-sure to alter eliminate boundary	pressure-swing distil-lation infeasible	azeotropic composition varies slightly with pressure (13 vol % water, 12.7 kPa; 12 vol % water, 101.3 kPa; 11.7 vol % water, 411 kPa)
6. exploit solid–liquid solubility differences	solution crystalliza-tion probably inferior	low (<0°C) temperatures required
7. exploit kinetic phenomena to cross boundary	mol sieve adsorption feasible with preconcentration	difference in kinetic dia (water 265 pm; IPA 420 pm) acceptable for 300 pm mol sieves; much water to adsorb; preconcentration step neces-sary for reasonable bed size
	membrane permeation feasible with precon-centration	dehydration of IPA known appli-cation of pervaporation, eg, cellulose acetate membranes; much water to pass through membrane; preconcentration step probably necessary to keep membrane area reasonable

^a Strategies from third section of Table 3, *Overcoming distillation boundaries*.

^b IPA = isopropyl alcohol.

^c To convert kPa to torr, multiply by 7.5.

For a binary system, the two possible opportunistic distillations are essentially identical and can be combined to concentrate the feed up to about 68° 2-propanol and produce pure water. The feed and products of the two possible strategic separations are ill-defined, whereas those of the opportunistic separation are known as outlined in Table 6.

Table 6. Feasible Opportunistic and Strategic Separations for IPA–Water System

Operation	Unit operation	Feed(s)	Products ^a
opportunistic separation from feed	fractionator	feed mixture, 40° IPA	D = IPA – water azeotrope B = water (target product)
strategic operations from feed	mol sieve adsorption	undefined IPA–water mix	water is adsorbed raffinate = mix enriched in IPA
	membrane permeation	undefined IPA–water mix	permeate = water raffinate = mix enriched in IPA

^a D = distillate; B = bottoms.

Because a preconcentration step is probably needed to make the final sequence more economical, it is logical to start with the opportunistic separation. This separation produces one of the products, pure water, as the underflow and a concentrated distillate appropriate for feed into either strategic separation. Arbitrarily choosing pervaporation first, the retentate has a composition on the 2-propanol-rich side of the azeotrope, whereas the permeate is pure water. No further strategic separations are required.

Addition of another opportunistic distillation to the flow sheet results in the production of the 2-propanol product (bottoms) and approximately the azeotrope composition out the top. The distillate is not a desired product and must be dealt with in some fashion. It is similar in composition to several other streams in the process, so recycle rather than further processing in additional equipment is considered. Referring to the Recycle heuristics, recycling directly back into the same column is unwise as water builds up. Recycle to the front of the process entails revaporization of this whole fraction with additional water and increases the membrane load. Recycle to the membrane inlet is probably best. Although the membrane load is increased, the fraction does not need to be distilled overhead again. The final dual-distillation, membrane permeation sequence is shown in Figure 4. A similar sequence can be constructed using molecular sieve adsorption as the strategic separation in place of pervaporation. Both sequences involve trade-offs between water removal in the membrane or adsorption bed, recycle rate, and 2-propanol production rate. Such an analysis is best carried out at a higher level of detail via process simulation.

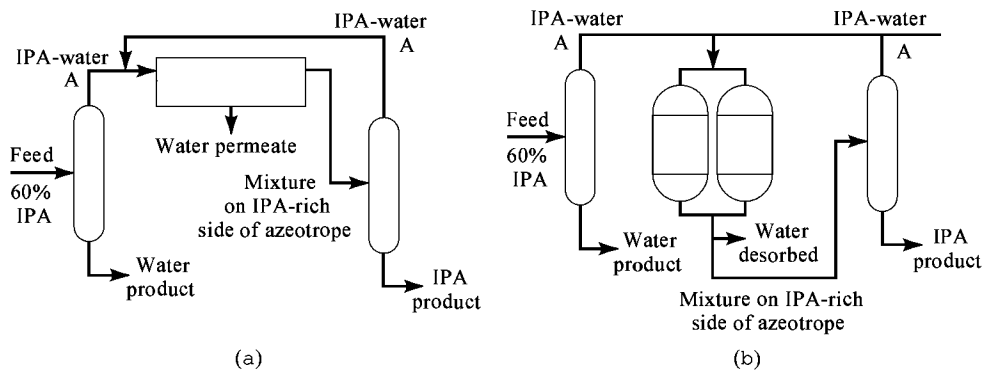


Fig. 4. Distillation of aqueous isopropyl alcohol (IPA), where A is the azeotrope, combined with (a) pervaporation and (b) adsorption.

Strategies Requiring Introduction of a Third Component. Although the original binary system does not exhibit liquid–liquid behavior, the first strategy for overcoming distillation boundaries suggests the addition of a mass separating agent to the system in order to cause the formation of a liquid–liquid region. A number of desirable characteristics for the MSA can be identified. The solvent must be partially miscible with either 2-propanol or water, or both. Hydrocarbons, ethers, higher ketones, and halogenated compounds are usually partially miscible with water. Moreover, the liquid–liquid region must be situated to allow for crossing the azeotrope or any distillation boundaries formed. Because the MSA must be recovered at some point in the process, it should have an appreciable boiling point difference from both water and 2-propanol. Further information on solvent selection for liquid–liquid extraction is available (24,25).

Different MSAs may lead to completely different separation systems designs. The systematic generation procedure given in the separations synthesis algorithm is demonstrated for two potential solvents, hexane and methylene chloride.

Water–2-Propanol–Hexane System. The residue curve map for the ternary system exhibits three distillation regions and a two-liquid phase region (Fig. 5). Each pure component is the high boiling node in a different distillation region. Thus, from the feed mixture a distillation boundary must be crossed in order to obtain pure isopropyl alcohol (IPA). The compositions of interest in this diagram include the feed mixture (60% water, 40% IPA), the two products (in different distillation regions), and potential MSA compositions such as pure hexane (high boiling node in region III), the ternary azeotrope (low boiling node in regions I, II, III), and points on the binodal curve, particularly tie-lines passing through azeotropes. Because the binary azeotropes are saddle points, and thus hard to regenerate by distillation, they are probably not good compositions to pick for an MSA. With the limited information available at this point, pure hexane is a reasonable choice as MSA. This is a decision point. Another MSA composition may lead to a different flow sheet. Moreover, it may be found that pure hexane is not required.

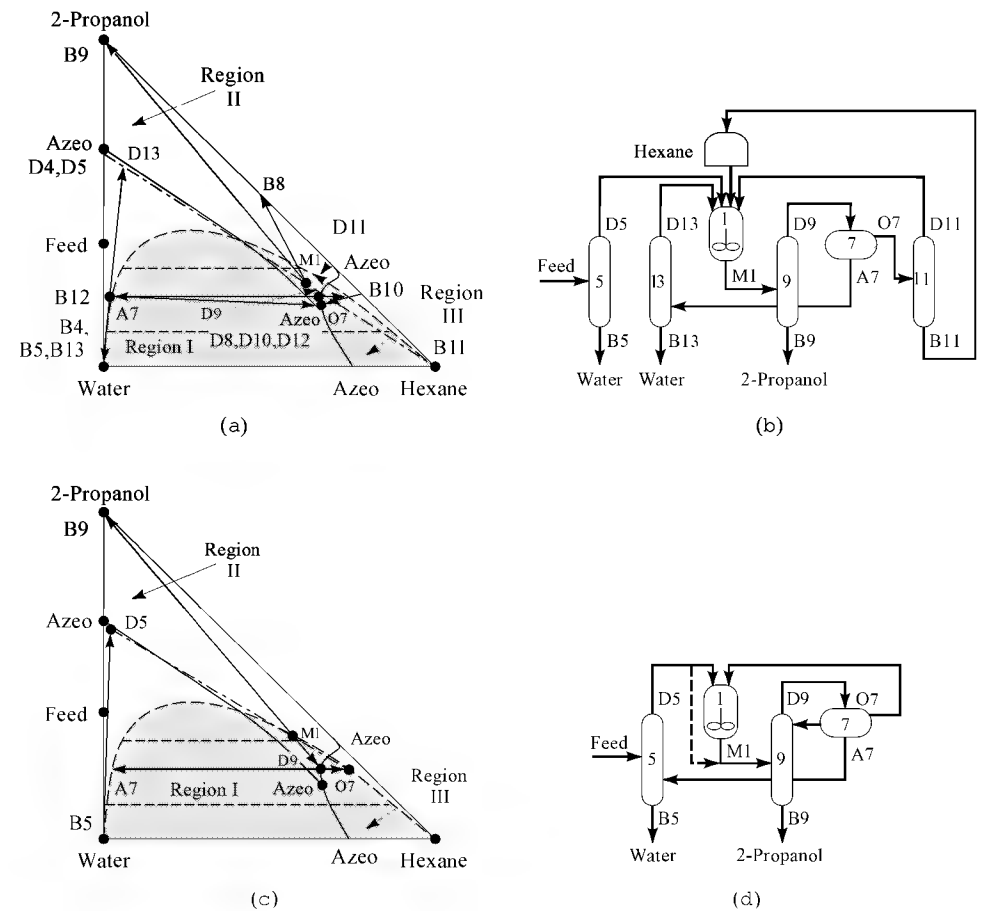


Fig. 5. 2-Propanol dehydration where Azeo is azeotrope, () are tie lines, and the shaded areas represent the region of two liquid phases. Initial dehydration (a) residue curve map and (b) flow sheet, and evolved 2-propanol dehydration (c) residue curve map and (d) flow sheet. The ~ in (b) and (d) represent a two-phase decanter. See text; see Table 7.

Without revisiting the kinetic-based methods already examined for the binary case, several strategic operations can be used to cross the distillation boundary between region I and II including Mixer 1, Extractor 2, and Decanter 3. The decantation strategy could be implemented either with a single feed, which must be in the liquid–liquid region, or by adding an MSA, which puts the overall mixture in the liquid–liquid region.

The feed compositions and products of each of these strategic separations remain ill-defined. The unspecified 2-propanol–water mixture, the input to each strategic separation, could be but is not necessarily the original feed composition. The MSA composition (pure hexane in this case) is such that one of the products of the strategic separation is in region II, ie, the strategic separation crosses the distillation boundary. Two opportunistic distillations from

the original feed mixture, Fractionators 4 and 5, are also possible. Because the feed is binary, these separations are identical and are the same as the preconcentrator previously discussed (Table 7).

Table 7. Operation for the IPA–Water System

Operation	Unit operation ^a	Feed(s)	Products
strategic operation for reaching region II	Mixer 1	undefined IPA–water mix, MSA	M1 = mixture in region II
	Extractor 2		E2 = extract in region II
opportunistic separation from feed		undefined IPA–water mix, MSA	R2 = raffinate in region I
		undefined IPA–water–MSA mix	O3 = organic phase in region II
			A3 = aqueous phase in region I
	Fractionator 4	mixture, 40% IPA	D4 = IPA–water azeo-trope
	Fractionator 5	mixture, 40% IPA	B4 = water D5 = IPA + water
strategic operation for reaching region III			B5 = water (target product)
	Mixer 6	undefined mixture in region II	M6 = mixture in region III
	Decanter 7	undefined mixture in region II	O7 = organic phase in region III
opportunistic separation from mixture M1			A7 = aqueous phase in region I
	Fractionator 8	M1	D8 = ternary azeotrope
			B8 = IPA + hexane or IPA + water
opportunistic separation from decant phase O7			
	Fractionator 9	M1	D9 = hexane + IPA + water
			B9 = IPA (target product)
	Fractionator 10	O7	D10 = ternary azeotrope
			B10 = hexane + IPA
opportunistic separation from decant phase A7	Fractionator 11	O7	D11 = hexane-IPA azeo + water
			B11 = hexane (target MSA)
opportunistic separation from decant phase A7	Fractionator 12	A7	D12 = ternary azeotrope
			B12 = water + IPA
	Fractionator 13	A7	D13 = IPA-water azeo + hexane
			B13 = water (target product)

^a See also Fig. 5.

Following step (4) of the algorithm and consulting the rules for Selecting Operations, the first operation should be Fractionator 5, an opportunistic separation that reaches a product. None of the strategic separations is able to reach a product composition. Application of Fractionator 5 to the feed produces the desired water product as underflow and a water–2-propanol mixture (D5) as distillate. Stream D5 requires further processing to obtain the 2-propanol product. At this point no other opportunistic separations can be applied, but stream D5 is a possible input to any of the three previously identified strategic operations. Arbitrarily picking the mixing strategy, sufficient hexane must be added to D5 to produce a composition in region II.

The strategy for boundary crossing has been implemented; however, by the addition of the hexane another critical feature has been created. Hexane must be regenerated, but it is in a different distillation region than the only remaining unprocessed stream (M1). In this case the possible boundary crossing strategic operations are Mixer 6 and Decanter 7. Two opportunistic distillations, Fractionators 8 and 9, can also be applied to M1 (decantation is also a possible opportunistic separation).

Again referring to the rules for Selecting Operations, the next separation should be Fractionator 9 which produces the 2-propanol product directly. If the distillation is driven far enough, stream D9 is in the two-phase region and is an appropriate input into the strategic separation Decant 7. Decanter 7 yields phases in region I and region III. Alternatively, D9 could be mixed with a sufficient amount of some stream (unknown at this time) in region III to bring the overall mixture into region III. Either strategy could be selected, but the decant is probably better as it can be accomplished without the addition of another stream. The organic phase of the decanter, O7, is in region III and needs to be further purified to obtain pure hexane. Two opportunistic separations, Fractionators 10 and 11, are possible.

Selection of Fractionator 11 gives pure hexane, which can be recycled to Mixer 1. The distillate D11, however, is a problem. It cannot be distilled because of its location next to a distillation boundary. It is outside of the two-phase region, so it cannot be decanted. In essence, no further separations are possible. However, using the Recycle heuristics, it can be mixed into the MSA recycle stream without changing the operation of Mixer 1 appreciably. However, as both outlet streams are mixed together, Fractionator 11 is not really needed. The mixture of hexane and isopropanol, O7, could have been used as the MSA composition in the first place.

The final loose end in the process is the aqueous decanter product, A7. The hexane must be removed before the mixture can be sent to wastewater treatment, ie, accepted as a water by-product. Two opportunistic separations, Fractionators 12 and 13, are possible. Selection of Fractionator 13 gives pure water underflow, and a distillate similar to D5. Distillate D13 can be recycled back and mixed with D5 without affecting the operation of Mixer 1. All streams are processed and the flow sheet produces both desired products (Fig. 5b).

Although, the flow sheet is complete, it does contain some redundant separation which adds to the capital cost. With the basic structure defined, the next step is to consider some evolutionary improvements. First, Fractionator 13 performs essentially the same job as Fractionator 5, ie, concentrating a mixture of water and 2-propanol. Fractionator 13 can be taken out and A7 can be recycled directly to the feed of Fractionator 5. Although the flow rate to Fractionator 5 increases, its basic separation function does not change. As the bubble point of A7 is somewhat higher than that of Feed F, A7 is probably best introduced as a separate feed lower in the column. Second, using pure hexane as the MSA was unnecessary. Fractionator 11 can be eliminated and O7 used as the MSA composition. Finally, considering the bubblepoint and amount of O7, operations Mixer 1, Fractionator 9, and Decanter 7 can be rearranged as a distillation–decanter combination with identical overall material balance lines and function. The evolved flow sheet is shown in Figure 5c and 5d. Other sequences can be generated by choosing different alternatives at the various decision points of the separations synthesis algorithm.

Methylene Chloride Alternatives. When the solvent is changed to methylene chloride the residue curve map contains two distillation regions and a liquid–liquid region (Fig. 3b). Pure methylene chloride is a saddle in distillation region II and therefore is probably not a good choice for the MSA composition. Rather, the compositions of interest are the water–methylene chloride azeotrope (high boiling node) points on the binodal curve, and 2-propanol-free mixtures of methylene chloride and water with compositions between those of the binodal curve and the methylene chloride–water azeotrope. The strategic operations are the same as with the hexane case (decant, extract, or mix) and again, two opportunistic fractionations are possible (preconcentration distillations).

In one possible sequence the MSA composition is chosen as water-saturated methylene chloride expected to be regenerated by decantation. The boundary-crossing strategic operation is to mix the feed with the MSA. The resulting two-phase mixture is opportunistically fractionated to produce the 2-propanol product as bottoms, and a mixture of water–methylene chloride as distillate. This distillate is opportunistically decanted to recover water-saturated methylene chloride MSA for recycle. The aqueous decanter phase is the water product, which optionally may be further purified by

stripping (Fig. 6).

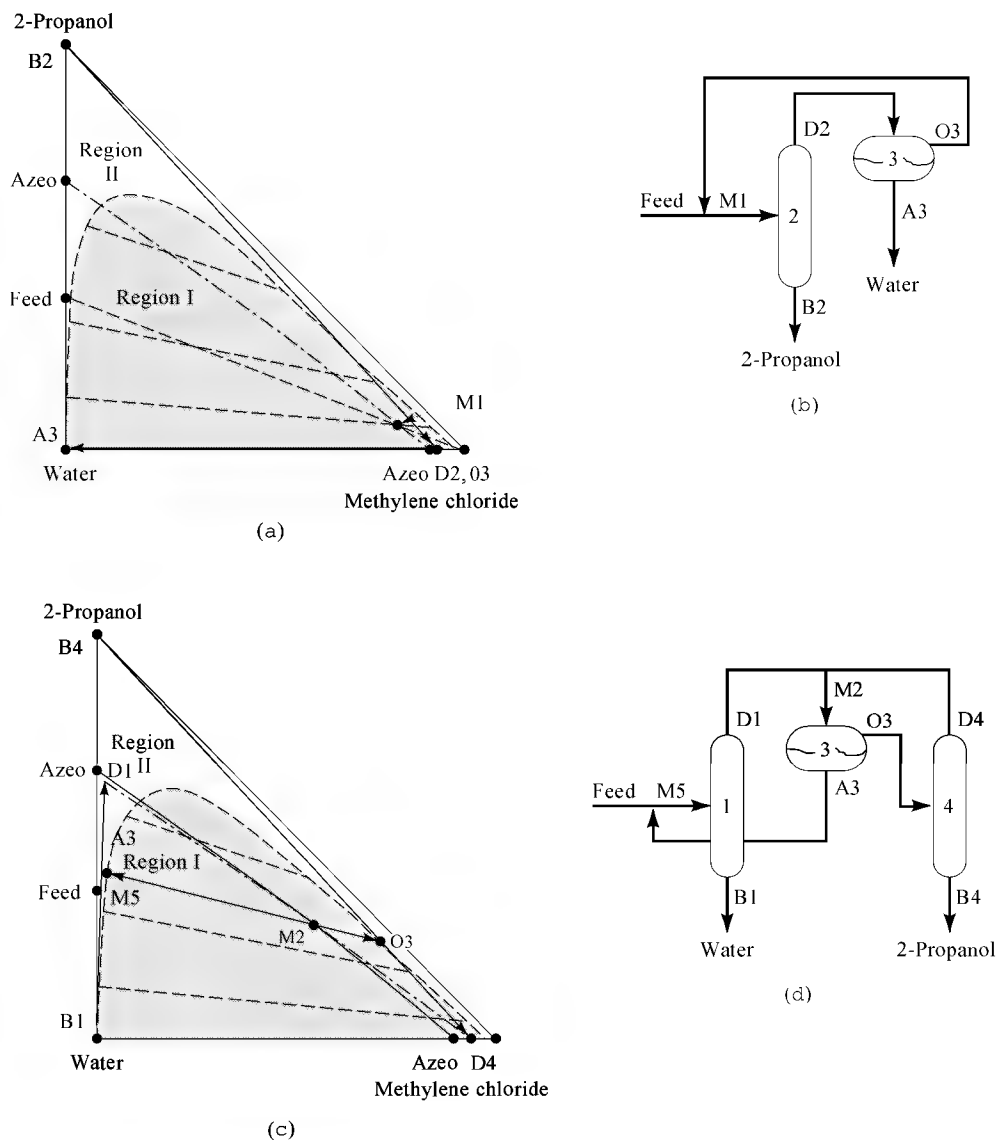


Fig. 6. 2-Propanol dehydration using methylene chloride where Azeo = azeotrope, the shaded areas represent the region of two liquid phases, (—) are tie lines, and ~, two-phase decanters: (a) residue curve map and (b) flow sheet, and alternative methylene chloride (c) residue curve map and (d) flow sheet. See text.

If an opportunistic preconcentration of the feed is used instead, an entirely different flow sheet results. In this case the MSA composition is a two-phase mixture of methylene chloride and water. Detailed simulations are required to determine which of these (or other) 2-propanol dehydration flow sheet alternatives is the economically advantaged process.

Flow Sheet Generation for Separating Gas–Vapor Mixtures

Whereas liquid separation method selection is clearly biased toward simple distillation, no such dominant method exists for gas separation. Several methods can often compete favorably. Moreover, the appropriateness of a given method depends to a large extent on specific process requirements, such as the quantity and extent of the desired separation. The situation contrasts markedly with liquid mixtures in which the applicability of the predominant distillation-based separation methods is relatively insensitive to scale or purity requirements. The lack of convenient problem representation techniques is another complication. Many of the gas–vapor separation methods are kinetically controlled and do not lend themselves to graphical-phase equilibrium representations. In addition, many of these methods require the use of some type of mass separation agent and performance varies widely depending on the particular MSA chosen.

No detailed parametric studies on the sequencing of gas–vapor separations have appeared in the process synthesis literature. However, several of the general separation sequencing heuristics formalized in previous studies are applicable to gas separations. The sequencing heuristics as modified for gas separations are (1) remove corrosive and hazardous materials first; (2) remove troublesome trace impurities first; (3) favor separations that match the desired products directly, ie, if a separation results in substream which requires no further separation and is a desired product, and if that product is the most plentiful in the mixture, remove it next; and (4) favor separations that give equimolar splits, ie, when ease of separation and compositions are similar, perform the separation that divides the feed as equally as possible. As for distillation sequencing for nonazeotropic liquid mixtures, a purely opportunistic strategy working from the original feed stream is adopted for gas–vapor separation scheme synthesis.

Gas-phase separations may be classified as enrichment, sharp, or purification separation, depending on the purity, recovery, and magnitude of the pertinent separation. The classification system allows for a certain amount of synergy, as several separation methods may be combined in order to achieve the desired result. Certain separation methods are favored for each category (26).

An enrichment is defined as a separation process that results in the increase in concentration of one or more species in one product stream and the depletion of the same species in the other product stream. Neither high purity nor high recovery of any components is achieved. Gas enrichment can be accomplished with a wide variety of separation methods including, for example, physical absorption, molecular sieve adsorption, equilibrium adsorption, cryogenic distillation, condensation, and membrane permeation.

A sharp separation results in two high purity, high recovery product streams. No restrictions are placed on the mole fractions of the components to be separated. A separation is considered to be sharp if the ratio of flow rates of a key component in the two products is >10 . The separation methods that can potentially obtain a sharp separation in a single step are physical absorption, molecular sieve adsorption, equilibrium adsorption, and cryogenic distillation. Chemical absorption is often used to achieve sharp separations, but is generally limited to situations in which the components to be removed are present in low concentrations.

The special case involving the removal of a low (2–3 mol %) mole fraction impurity at high (>99 mol%) recovery is called purification separation. Purification separation typically results in one product of very high purity. It may or may not be desirable to recover the impurity in the other product. The separation methods applicable to purification separation include equilibrium adsorption, molecular sieve adsorption, chemical absorption, and catalytic conversion. Physical absorption is not included in this list as this method typically cannot achieve extremely high purities. Table 8 presents a list of the gas–vapor separation methods with their corresponding characteristic properties. The considerations for gas–vapor methods are as follows (26–44).

Cryogenic distillation

Remove components with high melting points or those that may freeze at cryogenic conditions before processing.

Process probably not competitive for small-scale operations (<10–20 t/d product gas).

Process probably competitive if α (relative volatility) >2.0.

Physical absorption

Favor when partial pressure in feed of components to be removed is high (>1%).

Consider for enrichment operations when selectivity between key component is >3.

Consider for sharper separations when selectivity between key components is >4.

If composition of absorbate is to be reduced below ~100 ppm in lean gas product, consider following by chemical absorption or possibly adsorption.

Favor glycol absorption for large-scale desiccation operations requiring dew point depressions of <28°C (50°F); initial and operating costs of high volume glycol absorbers are typically much lower for small–medium dew point depressions than the corresponding costs of solid-phase desiccation.

MSA selection important.

Chemical absorption

Impurities must contain acid–base functional groups or other reactive groups.

Bulk stream must contain different acid–base or groups with different reactivities than impurities.

Favor when partial pressure in feed of components to be removed is low (<1%) and when desired removal is high (purities at the ppm level are not uncommon).

MSA selection is critical.

Membrane permeation

Not good for services where solids, tars, acidic gases are expected; stream should be condensed once before coming in contact with membrane to remove low boilers if possible.

More economically competitive if ideal permselectivity is >15 (highly dependent on membrane selection); indication of feasibility obtained with information on critical temperature and van der Waals volume.

Generally only good for enrichment separations with one- or two-stage units.

Cascade with recycle required to obtain one (relatively) pure permeate and enriched retentate.

Catalytic conversion

Favor catalytic conversion when impurities can be converted into desired product; further purification and/or separation steps may be unnecessary.

Because of destructive nature, catalytic conversion can be eliminated from further consideration if impurities are desired product.

Catalytic conversion can reach levels of <1–10 ppm of original contaminant.

Do not use combustion if any components to be kept in original form.

Catalytic combustion is feasible for purification processes only when impurities are at concentrations <10% of lower flammability limit and when bulk stream already consists of oxidation products, eg, airstreams, off-gases, and other inerts.

Catalytic oxidation should not be used when process stream contains halogenated organics (reaction products are often as objectionable as original impurity).

Hydrogenation sometimes useful to convert impurities to more easily removed or less objectionable species.

Adsorption (general)

Favor adsorption for processes that require essentially complete removal of water vapor (adsorptive dehydration is capable of achieving dew point depressions <45°C (80°F); molecular sieves are favored adsorbents.

Favor adsorption for small-scale desiccation operations; solid-phase desiccant systems are relatively simple to design and operate (generally are lowest cost alternative for processing small quantities of gas).

If objective is to recover adsorbed components (free of water vapor), inlet gas stream should be dried before molecular sieve adsorption process occurs (water adsorption on mol sieves is particularly strong because of polarity of surface).

When possible, adsorbent-fouling and -damaging components should be removed upstream of adsorber inlet.

High boiling organics (normal bp >150–180°C) preferentially adsorbed; extremely difficult to remove during regeneration cycle.

Under favorable conditions, low molecular weight organics may polymerize on surface of adsorbent (dialkenes, 1-alkenes, alkynes, conjugated double-bond systems, and epoxides are especially susceptible to this behavior).

Highly acidic/alkaline moieties may also cause permanent chemical alterations in the adsorbent (aluminas are sensitive to acid solutions, silica gels are sensitive to alkaline solutions).

Adsorption (enrichment or sharp separations)

Adsorption generally gives one highly pure product (adsorbate) and one product depleted in adsorbate.

For bulk separations, more highly adsorbed component should be in minority in feed, generally <10–20%.

Some large-scale enrichments (adsorbate consisting of >10–20% of feed); examples include hydrogen recovery, methane enrichment, and oxygen enrichment from air.

For reasonably sharp separations, ratios of equilibrium loadings of key components should be <2 (as ratio decreases enrichment of either product stream falls off considerably).

Gaseous species can only be separated by mol sieving effects when kinetic diameters fall into different zeolite aperture size categories (standard mol sieve diameters nominally 300, 400, 500, 800, 1000, 1300 pm).

Adsorption (purification separations)

Only for dilute solutions (<1–3% of feed), unless mol sieving effect also present.

Activated carbon adsorption generally uneconomical for removal of >1000 ppm contaminant from large stream unless able to regenerate bed; steaming is often easiest regeneration method but creates new wastewater problem (usually 3–5 kg steam required per kg of carbon for regeneration).

Do not regenerate molecular sieves by steaming; water typically very strongly adsorbed and may not be easily displaced by adsorbent in next absorption cycle.

Resin adsorbents (macroreticular polymer resins) generally good for removal of up to 1–2% of stream (often regenerable).

Usually best choice for desiccation of gases (<3% water) such as argon, helium, hydrogen, chlorine, hydrogen chloride, sulfur dioxide, ammonia, air, and chemical classes such as aliphatics, aromatics, halogenated compounds, oxygenated compounds (silica gel, zeolites, activated alumina all alternatives; some regenerable, some not).

Condensation

Enrichment operation, α (relative volatility) >7, required to get much separation because operation is far from equilibrium stage.

Favor condensation (a simple and cheap unit operation) for removal of low boilers from noncondensable gases when cooling water can be used as the condensing medium.

Table 8. Characteristic Properties for Gas–Vapor Separation Methods

Separation method	Characteristic properties	Separation principle	Separation agent	Industrial application ^a
cryogenic distillation	relative volatility	difference in volatility	heat transfer	separation of air into pure oxygen and nitrogen
physical absorption	selectivity in absorbent; chemical families	difference in volatility as altered by MSA	nonvolatile liquid solvent	removal of VOCs from exhaust air
chemical absorption	selectivity in absorbent	difference in volatility and related azeo-trope formation	nonvolatile re-active liquid solvent	removal of carbon dioxide and hydrogen sul-fide from crude natural gas
catalytic conversion	chemical family	conversion to different spe-cies through chemical reac-tion	chemical reaction	control of VOC levels in ex-haust gases
membrane permeation	ideal permselectivity, critical temperature, van der Waals volume	forced flow through semi-permeable membrane	membrane	separation of air into enriched oxygen and ni-trogen streams
mol sieve adsorption	kinetic diameter	exclusion of species from	solid adsorbent, possibly aided by temperature or pressure gradient	separation of air into pure oxy-gen and nitro-gen
equilibrium-limited ad-sorption	equilibrium loading	difference in affinity for adsorption on solid surface	solid adsorbent, possibly aided by temperature or pressure gradient	removal of water vapor from refrigerants
condensation	relative volatility	difference in volatility	heat transfer	removal of VOCs from exhaust air

^a VOC = volatile organic compound.

Flow Sheet Generation for Dissolved Solids

For systems of dissolved solids, such as inorganic salts in water or essentially nonvolatile or high melting organics, crystallization (qv) is the predominant separation method. VLE-based separation methods such as distillation are largely relegated to an auxiliary role, functioning primarily for solvent removal or solvent recovery from mother liquors. However, as for VLE-based methods, the application of crystallization is constrained by certain critical features characteristic of the solid–liquid equilibrium (SLE) of the system. Three of the more common critical features for dissolved solids systems are (1) simple eutectics, ie, in a system exhibiting a eutectic, pure crystals of one component are deposited as the solution is cooled; at the eutectic point, a solid mixture of fixed composition is formed and no further separation is achieved (a eutectic is somewhat analogous to an azeotrope in that it limits recovery and product purity); (2) solid solutions, ie, upon cooling, a multicomponent solid solution is deposited, but unlike a eutectic system, crystals of one component are never deposited in a pure state; analogously to a VLE pinch, solid solution formation does preclude the use of crystallization, but high product purity can only be obtained by multistage crystallization; and (3) compound formation, ie, the dissolved solid and solvent form one or more intermolecular compounds, but the solid cannot be obtained in a completely solvent-free form without further treatment (in aqueous solutions these are called hydrates; in nonaqueous systems, solvates). The particular compound formed is often a function of concentration and temperature. Solid–liquid-phase behavior is quite diverse and in many systems can be quite complicated. A review of solid–liquid-phase behavior is available (22). Methodologies for the generation of crystallization-based flow sheets for separation of multicomponent systems exhibiting eutectics, solid solution behavior, and compound formation have been presented (45–48). All methods make extensive use of phase diagrams for problem representation and analysis.

Crystallization of dissolved solids is induced commonly by one of four mechanisms: cooling, evaporation of solvent, reaction, and dilution. Cooling is favored for systems in which solubility shows a large decrease with decreasing temperature. Evaporation of solvent (generally isothermally) is the preferred mode of crystallization when solubility shows little or no temperature dependence or increasing solubility with increasing temperature. Adiabatic evaporation, simultaneous evaporation of solvent with induced cooling, sometimes called vacuum cooling, is favored when the temperature dependency of the solubility is moderate.

Dilution or reaction-induced crystallization often are viable alternatives for systems with high solubility, in which high recovery would require excessive evaporation or low temperature cooling. These techniques require the selection and addition of a mass separating agent (49,50). In dilution by drown-out a miscible nonsolvent is added which reduces the solubility of the solute. Another form of dilution, salting-out, requires the addition of a second more soluble solid which reduces the effective solubility of the original solute, ie, the common-ion effect. The addition of a component that reacts with the solute to form a less soluble product is another option, ie, acid–base neutralization with precipitation of the insoluble salt product. In general neither the salting-out agent nor the added reactant is easily recoverable. Development of the recovery scheme for the drown-out agent may be accomplished using the flow sheeting techniques discussed for liquid mixtures.

Generally adiabatic cooling by evaporation of solvent is favored over heat removal (evaporating cooling) through vessel or exchanger walls. Heat-transfer coefficients for condensation are generally larger than for conduction, reducing surface area requirements. In addition, surface cooling and evaporation tend to produce high supersaturation at the heat-exchanger surface, resulting in surface fouling and poor heat transfer. This is reduced by adiabatic evaporation. Theoretically complete recovery is possible by evaporation if the solubility of an impurity is not exceeded. Dilution or reaction schemes are often more complicated due to the addition of a MSA. With these items in mind, the order of preference for crystallization modes is adiabatic evaporation, evaporation, cooling, and dilution or reaction. Further considerations for the selection of crystallization methods were given previously.

BIBLIOGRAPHY

1.

I. E. Grossmann, in J. J. Siirola, I. E. Grossmann, and G. Stephanopoulos, eds., *Foundations of Computer-Aided Process Design*, Elsevier, New York, 1989, p. 105.

2.

S. D. Barnicki and J. J. Siirola, in J. D. Seader, ed., *Perry's Chemical Engineering Handbook*, 7th ed., McGraw-Hill Publishing Co., New York, 1996, section 13.

3.

J. M. Douglas, *Conceptual Design of Chemical Processes*, McGraw-Hill Publishing Co., New York, 1988.

4.

D. F. Rudd, G. J. Powers, and J. J. Siirola, *Process Synthesis*, Prentice-Hall, Inc., Englewood Cliffs, N. J., 1973.

5.

Y. A. Liu, in Y. A. Liu, H. A. McGee, Jr., and W. R. Epperly, eds., *Recent Developments in Chemical Process and Plant Design*, John Wiley & Sons, Inc., New York, 1987, section 6.

6.

N. Nishida, G. Stephanopoulos, and A. W. Westerberg, *AIChE J.* **27**, 321 (1981).

7.

R. M. Kelly, in R. W. Rousseau, ed., *Handbook of Separation Process Technology*, Wiley-Interscience, New York, 1987, section 4.

8.

V. M. Nadgir and Y. A. Liu, *AIChE J.* **29**, 926 (1983).

9.

Y. W. Huang and L. T. Fan, in M. L. Mavrouniotis, ed., *Artificial Intelligence in Process Engineering*, Academic Press, Boston, Mass., 1990, section 10.

10. Y. A. Liu, T. E. Quantrille, and S. H. Cheng, *Ind. Eng. Chem. Res.* **29**, 2227 (1990).

11. R. R. Wehe and A. W. Westerberg, *Chem. Eng. Sci.* **45**, 1 (1990).

12. D. B. Van Dongen and M. F. Doherty, *Ind. Eng. Chem. Fundam.* **24**, 454 (1985).

13. E. R. Foucher, M. F. Doherty, and M. F. Malone, *Ind. Eng. Chem. Res.* **30**, 760 (1991).

14. O. M. Wahnschafft, J. Koehler, E. Blass, and A. W. Westerberg, *Ind. Eng. Chem. Res.* **31**, 2345 (1992).

15. J. Gmehling, J. Menke, K. Fischer, and J. Krafczyk, *Azeotropic Data*, Parts I–II, VCH, New York, 1994.

16. S. D. Barnicki and J. R. Fair, *Ind. Eng. Chem. Res.* **29**, 421 (1991).

17. P. C. Wankat, *Ind. Eng. Chem. Res.* **32**, 894 (1993).

18. A. Eckles, P. Benz, and S. Fine, *Chem. Eng.* **98**, 201 (1991).

19. A. Eckles and P. Benz, *Chem. Eng.* **99**, 78 (1992).

20. J. W. Mullin, *Crystallization*, Butterworth-Heinemann, Boston, Mass., 1993.

21. C. G. Moyers and R. W. Rousseau, in Ref. 7, section 11.

22. N. S. Tavaré, *Industrial Crystallization*, Plenum Press, New York, 1995.

23. Y. L. Hwang, G. E. Keller, and J. D. Olson, *Ind. Eng. Chem. Res.* **31**, 1753 (1992).

24. T. C. Lo, M. H. I. Baird, and C. Hanson, *Handbook of Solvent Extraction*, Krieger Publishing Co., Malabar, Fla., 1991.

25. L. A. Robbins, *Chem. Eng. Prog.* **76**(10), 58 (1980).

26. S. D. Barnicki and J. R. Fair, *Ind. Eng. Chem. Res.* **31**, 1679 (1992).

27. K. D. Timmerhaus and T. M. Flynn, *Cryogenic Process Engineering*, Plenum Press, New York, 1989.

28. H. Springmann, *Chem. Eng.* **92**, 59 (May 13, 1985).

29. A. Kohl and F. Riesenfeld, *Gas Purification*, 4th ed., Gulf Publishing, Houston, Tex., 1985.

30. R. N. Tennyson and R. P. Schaaf, *Oil & Gas J.*, 78 (Jan. 10, 1977).

31. G. Astarita, D. W. Savage, and A. Bisio, *Gas Treating With Chemical Solvent*, John Wiley & Sons, Inc., New York, 1983.

32. C. England, *Chem. Eng.* **93**, 63 (Apr. 28, 1986).

33. T. L. Ho, *Hard and Soft Acids and Bases Principle in Organic Chemistry*, Academic Press, Inc., New York, 1977.

34. J. E. Hogsett and W. H. Mazur, *Hydrocarbon Processing*, 52 (Aug. 1983).

35. D. J. Stookey, C. J. Patton, and G. L. Malcolm, *Chem. Eng. Prog.*, 36 (1986).

36. R. McInnes, S. Jelinek, and V. Putsche, *Chem. Eng.*, 108 (Sept. 1990).

37. R. E. Reitmeier and H. W. Fleming, *Chem. Eng. Prog.* **54**(12), 48 (1958).

38. P. N. Rylander, *Hydrogenation Methods*, Academic Press, Inc., New York, 1985.

39. "Adsorption," in W. Gerhartz, ed., *Ullmann's Encyclopedia of Industrial Chemistry*, Part 9-6, Vol. B3, *Unit Operations*, VCH, Weinheim, Germany, 1988.

40. D. P. Valenzuela and A. L. Myers, *Adsorption Equilibrium Data Handbook*, Prentice-Hall, Inc., Englewood Cliffs, N.J., 1989.

41. P. C. Wankat, *Rate-Controlled Separations*, Elsevier Applied Science, New York, 1990.

42. R. T. Yang, *Gas Separation by Adsorption Processes*, Butterworth, Boston, Mass., 1987.

43. G. E. Keller, R. A. Anderson, and C. M. Yon, in Ref. 7, section 12.

44. J. J. Collins, *Chem. Eng. Prog.* **64**, 66 (1968).

45. S. Rajagopal, K. M. Ng, and J. M. Douglas, *AIChE J.* **37**, 437 (1991).

46. L. A. Cisternas and D. F. Rudd, *Ind. Eng. Chem. Res.* **32**, 1993 (1993).

47. S. R. Dye and K. M. Ng, *AIChE J.* **41**, 1456 (1995).

48. *Ibid.*, p. 2427.

49. T. J. Bruno and P. D. N. Svoronos, *CRC Handbook of Basic Tables for Chemical Analysis*, CRC Press, Inc., Boca Raton, Fla., 1989.

50. K. K. Nass, *Ind. Eng. Chem. Res.* **33**, 1580 (1994).

Scott D. Barnicki
Jeffrey J. Siirola
Eastman Chemical Company

SEROLOGY, HUMAN.

See FORENSIC CHEMISTRY

SHALE OIL.

See OIL SHALE

SHAPE-MEMORY ALLOYS

Metallurgical Basis for the Shape-Memory Effect

Shape memory or the shape-memory effect (SME) is a phenomenon associated with the martensite transformation, a first-order displacive transformation usually related to the hardening of steel (qv). In steel an alloy heated to the temperature where the elevated temperature face-centered cubic (fcc) austenite phase is stable, is rapidly cooled to produce the hard martensite phase. In certain alloys, however, the martensite transformation is thermoelastic, ie, the martensite forms and disappears on heating and cooling over a relatively small temperature range. The intermetallic phase in these alloys undergoes a displacive, shear-like transformation when cooled below a critical temperature designated as M_S . Upon further cooling, to a temperature designated as M_F , the transformation is complete and the alloy is said to be in its martensitic state. When this martensite is deformed, it undergoes a strain that is completely recovered when the alloy is heated. The recovery process starts at the temperature A_S and is completed at a higher temperature, A_F . Because this is a first-order phase transformation, there is hysteresis associated with the formation of martensite and its reverse transformation to the elevated temperature parent phase, which is usually also referred to as austenite. The temperatures M_S , M_F , A_S , and A_F depend on the particular alloy, alloy composition, and processing. The hysteresis loop associated with the transformation in a typical shape-memory alloy is illustrated in Figure 1.

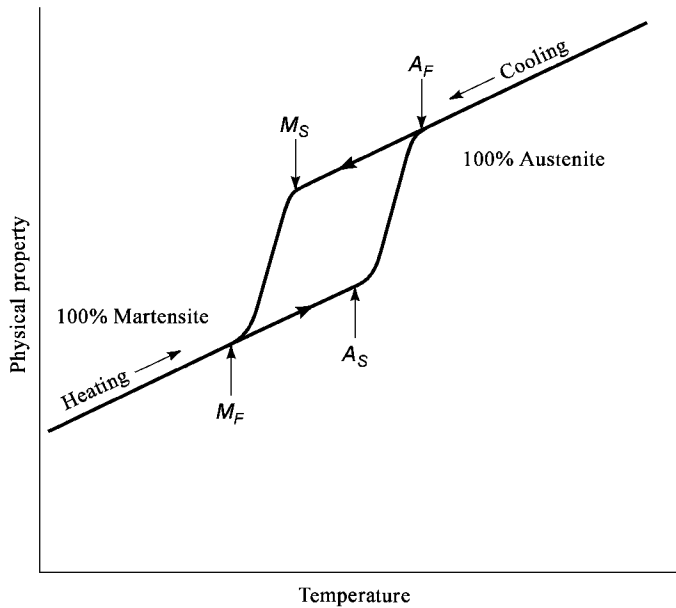


Fig. 1. Schematic of the hysteresis loop associated with a shape-memory alloy transformation, where M_S and M_F correspond to the martensite start and finish temperatures, respectively, and A_S and A_F correspond to the start and finish of the reverse transformation of martensite, respectively. The physical property can be volume, length, electrical resistance, etc. On cooling the body-centered cubic (bcc) austenite (parent) transforms to an ordered B2 or DO₃ phase and then to one of the various martensite structures.

The martensite in shape-memory alloys (SMAs) may also be isothermally generated by applying stress at a temperature above M_S . Because the martensite is unstable at this temperature, when the stress is removed the martensite disappears. Above a temperature designated as M_D , martensite cannot be generated, no matter how high the stress. The stress-induced martensite (SIM) gives rise to a mechanical type of shape memory called pseudoelasticity. These alloys exhibit, in addition to SME and SIM, another unique property, that of very high damping. Damping is the property which causes a vibration, once induced in a material, to decay. Bell bronzes have a low (<1%) damping, usually expressed as the specific damping capacity. A gray cast iron formulated for use as a machine-tool bed for resistance to vibration might have a specific damping capacity of 10⁰ %. By contrast, SMAs can have a damping capacity of greater than 40⁰ %. This characteristic can be exploited in smart or adaptive materials (see SMART MATERIALS (SUPPLEMENT)). Damping is the result of the high mobility of the interfaces between martensite variants or plates. When martensite is stressed, the deformation is not accommodated by the usual mechanisms that operate in conventional metals, such as a dislocation motion, slip, or grain boundary motion, but by the growth and shrinkage of martensite variants or the motion of twin boundaries. When subjected to an oscillating stress these mobile boundaries move back and forth, giving rise to a frictional loss that accounts for the damping.

The SME was first reported in 1951 involving a Au–Cd alloy (1). Many other alloys exhibiting this behavior have since been discovered; some of these are listed in Table 1. It was, however, the discovery of the SME in the nominally equiatomic Ni–Ti alloy that led to commercial applications. This

alloy is usually called nitinol, after Ni–Ti Naval Ordinance Laboratory (2). Nitinol and its variations are the most frequently employed shape-memory alloys, although two other SMAs, Cu–Zn–Al and Cu–Al–Ni, also find use in special applications. Nitinol is quite expensive, owing to the extremely tight composition control required to prepare an alloy having specific transformation temperatures. Moreover, nitinol is difficult to process into forms such as wire, rod, and ribbon that are required for commercial application. As a result, the copper-based SMAs were developed as less expensive alternatives. Although these latter alloys were employed for a period as actuators of various designs, shortcomings related to fatigue strength and thermodynamic stability restrict use. Certain ferrous alloys, listed in Table 2, have also been discovered. These are potentially less expensive SMA candidates, but as of 1996 have not achieved broad commercial use.

Table 1. Nonferrous SMAs^a

Alloy	Composition, ^b atomic %	Crystallographic structure change ^c	Hysteresis, Δ <i>T</i> , °C
Ag–Cd	44–49 Cd	B2–2H	15
Au–Cd	46.5–50 Cd	B2–2H	15
Cu–Zn	38.5–41.5 Zn	B2–9R	10
Cu–Zn–X	^d	B2(DO ₃)–9R, M9R, 18R, M18R	10
Cu–Al–Ni	28–29 Al, 3–4 Ni	DO ₃ –2H	35
Cu–Sn	15 Sn	DO ₃ –2H, 18R	
Cu–Au–Zn	23–28 Au, 45–47 Zn	Hesuler–18R	6
Ni–Al	36–38 Al	B2–3R	10
Ni–Ti	49–51 Ni	B2–rhombohedral; B2–monoclinic	1–2; 10–100
In–Tl ^e	18–23 Tl	fcc–fct	4
In–Cd ^e	4–5 Cd	fcc–fct	3
Mn–Cu ^e	5–35 Cu	fcc–fct	

^a Alloys are ordered except where noted.
^b To determine the percentage of the remaining constituents, subtract from 100^o %.
^c Fct = face-centered tetragonal.
^d A few atomic % of X = Si, Sn, Al, or Ga, plus 38.5–41.5 atomic % Zn.
^e Alloy is disordered.

Table 2. Ferrous SMAs Having Perfect or Near-Perfect SME

Alloy	Composition, ^a at %	Crystallographic structure change ^b	Ordering
<i>Small hysteresis</i>			
Fe–Pt	25Pt	LI ₂ –ordered bct	ordered
Fe–Pd	30Pd	fcc–fct	disordered
Fe–Ni–Co–Ti	33Ni–10Co–4Ti ^c	fcc–fct	disordered
<i>Large hysteresis</i>			
Fe–Ni–C	31Ni–0.4C ^c	fcc–bct	disordered
Fe–Mn–Si	30Mn–5Si ^c	fcc–hcp	disordered
Fe–Mn–Si + (Ni–Cr–Co)	15Mn–7Si–10Cr–10Ni–15Co ^c	fcc–hcp	disordered

^a To determine the percentage of the remaining constituents, subtract from 100^o %.
^b Bct = body-centered tetragonal; fct = face-centered tetragonal.
^c Value is in wt %.

SME involves a martensitic transformation from an ordered parent phase, yielding an ordered martensite that is crystallographically reversible and thermoelastic, although there are a few exceptions to the requirement of ordering for nonferrous alloys (Table 1), and more exceptions for the ferrous SMAs (Table 2). In some alloy systems the various martensites are internally faulted or internally twinned and can possess different crystal structures, the most common being 2H, 3R, 9R, 18R, monoclinic, and rhombohedral. In all known cases, however, an initial parent phase transforms to self-accommodating martensite plates, which are characterized by six plate groups, each plate group consisting of four variants. Owing to the self-accommodating nature of the transformation, the average shape deformation in a particular plate group is effectively zero. A body-centered cubic (bcc) parent phase that initially transforms to an ordered B2 or DO₃ phase is the dominant precursor or parent for SMAs. This is a result of the comparatively large thermodynamic difference that exists between the transformation from bcc to 2H, 3R, 9R, 18R, and monoclinic martensites and the typical ferrous martensites that transform from a face-centered cubic austenite to a bcc or body-centered tetragonal (bct) martensite.

The SME process can be illustrated by the Cu–Zn system, one of the first SMAs to be studied. A single orientation of the bcc β-phase on cooling goes through an ordering process to a B2 phase. In a disordered alloy, the lattice sites are randomly occupied by both types of atoms, but on ordering the species locate at particular atomic sites, yielding what is called a superlattice. When the B2 phase is cooled below the *M_s* it transforms to self-accommodating variants of martensite. The habit planes for the transformation are symmetrically disposed around the ⟨110⟩ family of planes, of which there are six in the cubic system. Thus, for the ⟨011⟩ plane the four variants are (2 11 12), (2 12 11), (2 12 11), and (2 11 12). These are termed the plate group. The six ⟨110⟩ planes of the bcc parent then yield a total of 24 martensite variants. When the martensite group is deformed, the deformation proceeds by the gradual conversion of four variants to a single martensite plate. The surviving orientation depends on the strain direction, and the growth or shrinkage of the variants is such as to minimize the strain energy accumulation. In other words, the surviving variant is the one that allows the greatest extension in the applied stress direction. This varies with the nature of the strain, tension, compression, bending, or torsion. Once a group becomes a single variant, it can change its orientation to that of an adjacent group by the process of twinning. Upon heating, the reverse transformation from martensite to the β-phase occurs between *A_s* and *A_f*, and the single crystal (plate) of martensite transforms to the single β-crystal having the original orientation of the

parent phase.

In the case of Cu–Zn and other systems that have relatively complex ordered martensite structures (9R or 18R), the reverse transformation is crystallographically restricted. This means that although there are many variants which can form on transformation from parent to martensite, only a single parent orientation is possible in the reverse, or shape-recovery, transformation. This process is illustrated in Figure 2a; the shape recovery effect is shown in Figure 2b. A detailed discussion of the crystallography of martensitic transformations and lattice-invariant shear is available (3,4). When a martensite group is deformed to coalesce into a single orientation, the dominant mechanism is twinning, where each twin is actually an alternate variant of the martensite crystal. Thus, for the four variants that cluster about the $\langle 110 \rangle$ habit plane, each orientation is a twin of another, and by the degenerate variant-twin relationship, a group of martensite plates from a single parent crystal can, on deformation, coalesce to a single crystal (single variant) of martensite. The parent phase, usually ordered B2 or DO₃, transforms to one of the martensite crystal forms that exhibit shape memory. Examples are 2H, Cu–Al–Ni, Cu–Sn, and Ag–Cd; 3R, Ni–Al; 9R, Cu–Zn; 18R, Cu–Zn–Al; and the monoclinic, NiTi.

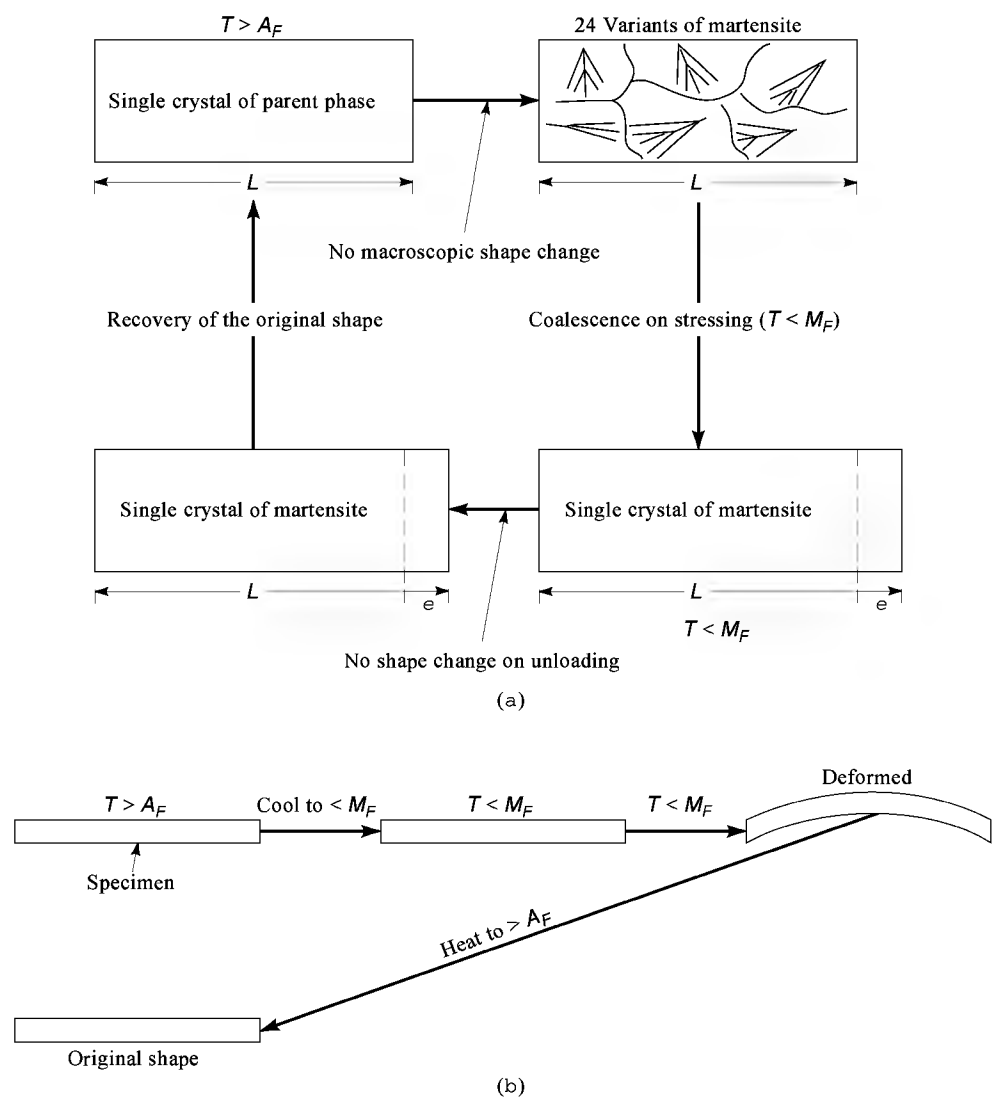


Fig. 2. The shape-memory process, where T is temperature. (a) The cycle where the parent phase undergoes a self-accommodating martensite transformation on cooling to the 24 variants of martensite. No macroscopic shape change occurs. The variants coalesce under stress to a single martensite variant, resulting in deformation. Then, upon heating, they revert back to the original austenite crystallographic orientation, and reverse transformation, undergoing complete recovery to complete the cycle. (b) Shape deformation. Strain recovery is typically ca 70%.

A distinction exists in the habit and deformation characteristics of the 2H and 3R types and the 9R and 18R martensites. The former are internally twinned and deformation occurs by a detwinning of a variant plate. The latter are internally faulted and deformation proceeds by variant-to-variant coalescence, followed by group-to-group coalescence. Although these structural differences exist, the self-accommodating habit-plane grouping with respect to a $\langle 011 \rangle$ plane is common to all systems exhibiting the SME. The 9R martensite is derived from a B2 parent; the 18R transforms from a DO₃ superlattice. The difference in stacking of $\langle 110 \rangle$ planes in these two structures results from the requirement for an invariant plane strain that involves a restricted stacking of close-packed planes. In order to obtain the required plane strain condition, both the 9R and 18R contain stacking faults to provide the necessary accommodation. The atomic displacements required to yield these structures combine the processes of shuffling and shear. The 3R martensite twin plane is identical to the 9R and 18R fault plane. For the case of 2H martensites, no such twin-fault correspondence exists, and the twin is derived from a different parent $\langle 110 \rangle$ plane. A theory of thermoelastic martensite deformation and shape recovery has been developed (5).

Simple shape-memory behavior, that is, one-way memory, may be extended to two-way memory. When a deformed martensite is heated to recover its original shape, it is a one-time event; to repeat the sequence the martensite must again be deformed. In two-way memory, the part spontaneously changes from one shape to another upon cooling or heating. This behavior, requiring a special conditioning of the martensite, is effected by limiting the number of martensite variants that form upon cooling through the application of an external stress during the transformation. The stress favors the initial formation of selected variants, similar to the stressing of a martensite group causing the selective growth and shrinkage of suitably oriented plates. Limiting the number of variants formed reduces the self-accommodating feature of the transformation and increases the residual internal stress. By repeating the cycle a number of times, referred to as the training process, the restricted variant group and its associated internal stress spontaneously revert to the parent phase on heating and then to a singular martensite on cooling.

Another property peculiar to SMAs is the ability under certain conditions to exhibit superelastic behavior, also given the name linear superelasticity. This is distinguished from the pseudoelastic behavior, SIM. Many of the martensitic alloys, when deformed well beyond the point where the initial single coalesced martensite has formed, exhibit a stress-induced martensite-to-martensite transformation. In this mode of deformation, strain recovery occurs through the release of stress, not by a temperature-induced phase change, and recoverable strains in excess of 15% have been observed. This behavior has been exploited for medical devices.

Applications of the SME

Applications for nonferrous shape-memory alloys fall into the following categories: one-way SMA devices, virtual two-way devices using one-way SMAs, two-way SMA applications, and SIM and superelastic devices. Ferrous alloy devices are considered separately. Over 10,000 patents have been issued on applications for SMAs; the actual number of these that have resulted in commercially successful products, however, is probably only about 100. Some of the limitations that impact the chances for broad commercial acceptance of these alloys are (1) recoverable strain, depending on the alloy, varying from 1 to 8°; (2) cyclic strain limit, ie, the required fatigue life determines the permissible shape recovery from 4 to <0.5%; (3) cyclic stability, ie, some alloys exhibit a change in transformation temperature after many cycles of shape recovery; (4) temperature limitations, ie, the maximum shape-recovery temperature known for an alloy requiring long-term exposure is about 130°C; (5) hysteresis, ie, alloys having a set hysteresis are required for various applications and the alloys available do not always meet the requirement; (6) processibility, ie, compositional control is exacting and the hot and cold working of these alloys is quite difficult compared to conventional alloys; and (7) fabricability, where the ease of forming a device or the available room temperature ductility of the alloy is not always satisfactory.

One-Way SMA Applications. The first successful application for nitinol was as a tube coupling for aircraft hydraulic control lines. In this application, a coupling is fabricated having an inner diameter smaller than the tube to be joined. The coupling is then cooled to its martensitic state and expanded by about 8° so that it fits over the tube. On warming to the shape-recovery temperature, the coupling shrinks and produces a joint having a very high reliability. The original couplings required cooling to cryogenic temperature and after expansion were maintained at this temperature until installation. A newer alloy, NiTi alloyed with Nb, has a very wide hysteresis that allows the alloy, after cryogenic expansion, to be stored at room temperature. Upon installation the coupling is heated to about 100°C for recovery, creating the joint. A primary consideration in the use of SMA couplings is the range of temperatures to which it is to be exposed. The high temperature is limited by stress relaxation and thermodynamic stability; the low temperature limit is dictated by the M_s temperature of the alloy, the temperature where loosening occurs. SMA couplings of this type have been installed in chemical and petroleum plants, power plants, and in piping systems (qv) aboard ships.

Using the wide hysteresis alloy, seals of various types have been developed based on the closing force of a wire ring of suitable diameter. Such seals are used for a wide variety of electronic device closures, eg, the attachment of shielding on electrical cables, and as closures on thin-wall cylindrical packages used in the semiconductor industry (see PACKAGING, ELECTRONIC MATERIALS). These seals avoid the heat of soldering, brazing, or welding (qv); thus, there is less chance of damage to the temperature-sensitive electronic components. Using couplings similar to the tube coupling, SMA fiber optic connectors have been produced that provide the necessary accuracy of axial alignment and a good hermetic seal, and do not require excessive heat for installation.

In many of the operations of spacecraft, it is necessary to separate the craft from its mooring before launch, and then to separate the booster and subsequent stages from the payload. The most frequently employed device for these procedures is an explosive bolt. These are also used in a variety of safety-release and closure systems in industrial plants. Because these devices are considered hazardous, the installation involves many precautions. In addition, a severe electrical storm can induce voltages in the system sufficient to activate the explosive charge accidentally. A SMA device has been developed that consists of a prestrained cylinder surrounding a bolt that has a machined notch in its surface (6). When the SMA device is electrically heated it expands, causing the bolt to fracture at the notch, releasing the load. The electrical energy required is well above that which can be generated by an accidental static electric charge pickup.

Virtual Two-Way Devices Using One-Way SMAs. Although alloys such as NiTi are considered to be one-way shape-memory alloys, by using a bias spring to provide the reverse motion (a recocking), a two-way device can be designed. A typical linear actuator consists of a shape-memory spring that produces motion and force in one direction and a steel spring that opposes this motion. When heated, the SMA spring has sufficient force to overcome the bias, but when the system cools, the bias ring has greater force and causes a reverse motion. A salient feature of SMA devices is that the modulus of elasticity increases with increasing temperature. This is quite unlike conventional materials. The modulus difference between the martensitic structure and the parent-phase structure is typically 300°. Virtual two-way devices have found many applications, for example, as electrical connectors (qv) in computer systems. At the high operating frequencies (>100 MHz) of modern computers there are problems of cross-talk between adjacent channels, noise, switching delays, and signal velocity, requiring very close pin spacing, shielding, and high contact force. The high forces needed in these multipin connectors make it difficult to make the male–female connection without damage to the contact surfaces. A device that opens and closes the jaws of the female component, thus providing what is termed a zero-insertion-force (ZIF) connector, is readily produced using SMA actuators. The closing force is provided by a SMA spring-like member when at ambient temperature. If the device is cooled by a cold jet of inert gas, its modulus decreases as it transforms to martensite and a second spring then opens the connector, allowing easy insertion of the male connector. When the connector warms back to the normal operating temperature, it closes with the required high force, providing a low contact resistance connection. Other versions of this device have been employed for military electronics requiring resistance to loosening under severe vibration conditions.

Circuit breakers have severe demands on performance and reliability, and the bimetals usually employed in the tripping mechanism have several disadvantages. Because their motion on heating is very small, the circuit breaker, depending on its size and capacity, must be calibrated to ensure that at the prescribed trip current the circuit breaker opens. In addition, because the force output of a bimetal is modest, a large circuit breaker may require the use of a cascading series of spring-actuated levers to provide the required interruption force and speed. One of the salient features of shape-memory actuators is that they produce larger motions and forces in a lighter and smaller device than a wax motor, solenoids, bimetals, or electric motors. Shape-memory actuators do, however, have one disadvantage, ie, a relatively low cyclic speed. Although a shape-memory device can be heated in milliseconds or less to provide the shape recovery, the reverse cycle of cooling is limited by convective, conductive, or radiation heat transfer, thus placing a practical limit on all but very small actuators having a cyclic rate of about 4 Hz.

Some notable examples exist of actuators that perform a function far more efficiently than alternatives, in some cases a function that cannot be performed by other actuator devices. An automatic transmission fluid temperature controller for Mercedes Benz cars is one example. Smooth shifting at low temperatures is assured by the control of the transmission fluid pressure, allowing full fluid flow only when the proper operating temperature is reached. Other automotive applications have been designed that take advantage of the compact, simple actuators that are possible using the high force and large strain capability of SMAs. Many devices have been demonstrated, eg, engine controls, climate controls, mirror motions, windshield wipers, cooling fan control, and retractable headlights. The best available NiTi SMAs have an upper actuation limit of about 80°C, and summer under-the-hood temperatures of cars operating in North America and in many other countries can reach over 125°C. Although the Cu–Al–Ni family of SMAs can be alloyed to give an A_s temperature in excess of 150°C, the long-term thermodynamic stability of these alloys is considerably lower. There is a considerable amount of research ongoing as of 1996 to develop high temperature shape-memory alloys, and many applications await their successful implementation. NiTi alloys having additions of Hf, Zr, or Pt have shown promise, and A_s temperatures in excess of 200°C showing good stability have been noted. Either cost or fabricability limitations have inhibited commercialization as of the mid-1990s.

Other successful actuators have been developed for various domestic safety devices. One such device is a shape-memory-actuated valve that fits in line with a bathroom sink, tub, or shower, and in the event that through some system imbalance or malfunction the water discharge approaches the temperature that would cause scalding, the valve automatically shuts off the flow and does not allow water flow to resume until the temperature is safe. A similar thermally triggered valve has been developed for shutting off the flow in industrial gas or fluid lines in the event of an excessive temperature or a fire. Shape-memory actuators are also candidate actuators to replace the glass bulb or fusible link in overhead fire extinguisher systems. Regulations stress the protection of human life, requiring that the sprinkler actuation speed be decreased to 14 s, well within the capability of an inexpensive SMA device but more difficult to achieve in the conventional trip devices.

In addition to the circuit breaker, there have been a number of other SMA applications for various functions in electric power generation (qv), distribution, and transmission systems. One such device is a thermal indicator that provides a signal visible from the ground of a hot junction or connector in a distribution yard. Such hot spots occur as a result of the loosening of bus bar connectors owing to cyclic temperature as the electric load varies. In addition to the use of SMA flags as a hot-spot indicators, actuators that automatically maintain the contact force in a bus bar connection have been demonstrated. Based on a Belleville washer fabricated from a Cu–Al–Ni SMA trained to exhibit two-way memory, these washers, when heated by a hot joint, increase their force output and correct the condition. A 30 mm diameter washer 3 mm thick can produce a force of over 4000 N. Similar in purpose

to circuit breakers, fuses are also candidates for shape-memory actuation, particularly on the large fuses used in power station distribution yards. The difficulty of detecting which phase in a system has blown has led to the design of resettable fuses as well as fuses which have shape-memory indicator flags that signal which phase has blown. The performance of overhead transmission lines is degraded when, as a result of overheating, usually from a combination of ambient conditions and overload, the lines sag to a point where ground capacitance becomes excessive. In addition, safety regulations require that the distance from the conductors to the ground be some minimum height. To correct the sag, shape-memory actuators in parallel with the conductors re-tension the lines when the lines reach a specified temperature. Sag-control models using SMA devices have been demonstrated by the electrical utilities in the Ukraine and Japan.

Another problem, prevalent in areas where severe icing conditions are met, is referred to as galloping of power lines. When ice forms on a power line, there is frequently a prevailing wind which causes the ice to take a teardrop or airfoil shape. This foil provides an aerodynamic lift to the conductors and under certain conditions the conductors can go into a resonant vibration such that large standing waves are created that exert enormous forces on the system. Miles of power lines and the towers along them have been destroyed by this phenomenon.

A system for exploiting shape-memory damping to reduce the severity of these resonance conditions has been designed and as of this writing (ca 1996) is being tested. Various schemes are being investigated for damping transmission line vibration using passive damping devices. In a related problem, the use of shape-memory alloys for seismic vibration control has been studied in national earthquake engineering centers in the United States and Japan. Two modes of operation have been explored: ground isolation and cross-brace stiffening. Ground isolation is effected by using large SIM devices in the foundations that exhibit a large change in modulus as a result of the stress developed by the seismic event. Copper-based SMAs have been studied for this application owing to availability in large cross-section and at moderate cost. In cross-brace stiffening of a building, a shape-memory element is placed in a diagonal position in the structure. In the event of a seismic disturbance the shape-memory element is electrically heated, producing an increase in its modulus and thus creating a stiffening effect to resist building deflection. In the first case the damping is passive and in the second a signal from an accelerometer is sent to a controller, which powers the SMA actuator. Seismic control is a problem worldwide for general civilian structures but is a particular concern in cases of power plants that are sited near a known geological fault. More details on the use of shape-memory alloys for structural vibration control are available (7,8).

Another problem in power generation is the cavitation erosion that occurs on the faces of hydroelectric turbines. This severe form of erosion must be periodically repaired to avoid a serious degradation in turbine efficiency. SMAs have a unique resistance to cavitation erosion and as of the mid-1990s exhibit orders of magnitude lower levels of metal loss than any other alloys. Early studies showed the exceptional cavitation resistance of NiTi. The reactive nature of NiTi has, however, impeded extending laboratory trials to the cladding of giant turbine runners. The CuAlNi SMA shows equivalent cavitation resistance and offers the advantage of being amenable to weld overlay techniques.

Space exploration presents a unique array of problems, involving in-space construction of structures, deployment of large antennas, control of antenna geometry, actuators for robotic devices, and control elements for a broad variety of scientific experimental apparatus. SMA joining techniques have been designed for the robotic assembly of large space structures. Usually, these structures are truss-like members having many struts that must be joined using a minimum of energy and a minimum of extravehicular work by the astronauts. SMA couplings which can be attached and then reattached to move to a new location are possible. These could be a useful tool for this area of space engineering.

One problem inherent in space structures (9) is vibration, which can impact on the function of space experiments and the accuracy of instruments. Large flexible structures, once put in motion by some external or internal force or shift of mass, have by themselves no substantial damping, operating in the vacuum of space. In essence, a smart material is one which can change its characteristics in such a way as to correct a condition which degrades its performance. These systems have incorporated into their structures sensors (qv), controllers, and actuators. The detector, sensing a vibration, sends a signal to a microprocessor-controller that in turn energizes some form of actuator to change the local structural dynamics, causing the vibration to decrease or be canceled. Shape-memory wires and ribbons embedded in the composite materials of construction have proved to be very effective. The wire or ribbon is prestrained before being embedded and then held under constraint while the composite, typically a graphite epoxy, is cured. When these SMA actuators are electrically heated, they attempt to recover their original dimensions. Owing to the constraint offered by the composite to which the SMAs are bonded, however, they can only produce an in-plane stress that alters the modal response of the structure, reducing its vibrational amplitude. In addition to acoustic and vibration control, SMA actuators are also used in smart materials for shape control, for example, changing the contour of an airfoil or hydrofoil, fine-tuning the contour of an antenna, and changing the focal point of microwave reflectors.

The design of smart materials and adaptive structures has required the development of constitutive equations that describe the temperature, stress, strain, and percentage of martensite volume transformation of a shape-memory alloy. These equations can be integrated with similar constitutive equations for composite materials to make possible the quantitative design of structures having embedded sensors and actuators for vibration control. The constitutive equations for one-dimensional systems as well as a three-dimensional representation have been developed (7).

Superelastic and Pseudoelastic Devices. Medical instruments and devices have become an important application for shape-memory alloys (see PROSTHETICS AND BIOMEDICAL DEVICES). Both shape recovery and superelastic behavior are exploited. The first application that has achieved broad acceptance is the use of pseudoelastic SMA wire for orthodontic arch wires (see DENTAL MATERIALS). These wires, attached to the teeth by small adhesively bonded clips, provide the correcting force for adjusting the position of teeth. The wires have a resistance to permanent set about 10 times that of the traditional stainless steel wires. This translates into much faster correction of the misalignment and far fewer periodic adjustments by the orthodontist. The use of traditional bridges and plates for teeth replacement is giving way to tooth implant procedures. This involves the attachment of a metal peg in the jaw bone and when the two are stabilized, the new tooth is attached to the peg. Using conventional implant techniques it can take over six months for the metal implant to stabilize before the replacement tooth is attached. Using shape-memory implant posts that lock into the bone immediately, reduces the time for the procedure to days. In maxillofacial surgery, which often involves reshaping the mandible and other cranial bones, shape-memory bone fasteners are proving superior to the usual screwed metal plates. If the SMA plate is prestrained and then attached to the two bone elements, as it warms to body temperature it pulls the mating surfaces into intimate contact, accelerating the healing process. Fractures in other bones such as in the arms and legs have benefitted from this procedure. Often fractures are associated with detachment of tendons that are normally difficult to reattach to the bone. Shape-memory anchors can make this a simple procedure. A small anchor with shape-memory wings is forced into a small hole drilled in the bone, and the wings provide secure locking in the hole. The tendon is then sutured to the anchor and in time attaches itself to the bone (10).

If it is not dissolved or trapped, an embolism moving from the lower extremities can be life-threatening. People afflicted with phlebitis are particularly susceptible to this problem. A shape-memory trap has been devised that, when deployed in the vena cava, is like a multileaved mesh that traps a traveling embolism, retaining it until medication can dissolve it. Introduced in a folded form by a catheter, the mesh is prevented from deploying by subjecting it to a flow of cold saline water. Once in place, it is released from the catheter and, warmed by body heat, opens into its final shape (11).

A variety of procedures have been developed to improve spinal fusion. Shape-memory rings are used to separate adjacent vertebrae and allow bone implants or bone chips to be inserted to develop the solid fusion section. Abnormal curvature of the spine, ie, scoliosis, has seen dramatic improvement in treatment in a procedure developed in China involving NiTi rods. The rods, cooled to their martensitic condition, are bent to match the curvature of the spine and, while maintained at a low temperature, are attached to vertebrae in a number of places along the spine. When the incision is closed, the SMA rods warm to their shape-recovery temperature and in attempting to revert to their original straight shape exert a continuous moderate corrective force on the spine. Badly curved spines have been brought to an essentially straight condition in about two years' time, and even after the rods are removed, there is little tendency to revert to the previous condition. Fracture of bones, common in other treatments, is avoided in this procedure (12).

NiTi superelastic wires have a remarkable resistance to kinking or developing a permanent set after bending. This feature is exploited in the long thin wires used to guide a catheter to its required position in the body. The wires allow safe introduction of a catheter through the sometimes tortuous passages of veins and arteries in cardiology procedures such as angioplasty. This flexibility and resistance to kinking of superelastic SMAs also makes possible superior needles and probes for arthroscopic procedures. Breast lesions, once located by radiology, are difficult to identify when the surgeon operates to remove them. A superelastic probe having a curved hook end can be introduced by a hypodermic needle during radiography and when the needle is retracted the curved end keeps the probe, called a Homer Mammalok (13), in place, thus identifying the location of the lesion. Stents, spiral tubular spring-like devices used to restore the patency of a vessel such as an artery, various ducts, and even the esophagus, have been made using conventional

plastics and metals, but with only partial success. The ability to make such a device of SMA ribbon or wire in coiled form, then stretch it almost straight to allow its placement in a catheter for introduction to the desired location, and then on deployment have the ribbon reform into a rigid coil that supports the vessel walls, is a great advance. This type of stent has also been demonstrated to be a fast and effective repair device for an abdominal aortic aneurysm. This latter defect, if left untreated, has a high incidence of mortality.

In addition to these medical applications, superelastic wires in very large quantities have been used for eyeglass frames, where their resistance to accidental damage and light weight are attractive features. Another large market is in the cellular telephone. The short antenna of superelastic shape-memory wire is very resistant to kinking and bending damage. A substantial tonnage market for superelastic memory alloy wire also exists for stiffening wire in brassieres. Resistance to permanent bending during machine washing and drying is provided by the SMA.

Devices required to produce exceptionally high forces have also been developed. Industrial force generators, using the recovery of a pretrained shape-memory alloy shape having the capability of producing forces from 10 to 100 t, have been introduced. In one form, developed in Japan, a number of small cylinders of NiTi are compressed and then placed in a cylindrical holder, which in turn is placed into a hole, usually in a cement or stone structure that is to be fractured. When the set of NiTi cylinders is electrically heated they go through shape recovery, expand, and fracture the structure, but without the hazard, noise, and dust created by conventional explosives (14) (see EXPLOSIVES AND PROPELLANTS). A Russian force generator has been developed that is capable of exerting the 100 t of force necessary to remove gas turbine wheels from their shafts.

Ferrous Shape-Memory Alloys. Ferrous shape-memory alloys have potential as low cost, readily fabricated one-way devices. These alloys, with one exception, do not form thermoelastic martensite. The alloy family of greatest interest, Fe–Mn–Si, forms martensite on cooling. However, if this martensite is deformed, it does not exhibit shape memory on heating to the A_F temperature. On the other hand, if the alloy is subjected to sufficient stress, an epsilon martensite forms by the SIM reaction. This deformation is recovered on heating, yielding a one-way shape-memory behavior. The basic alloy does not have the corrosion resistance required for many industrial applications, and to improve this feature, the alloy has been further alloyed with Ni, Cr, and Co (see Table 2). The recoverable strain is modest when the alloy is subjected to its first deformation and recovery cycle, but on subsequent cycling, more or less similar to the training cycles used to impart two-way memory, the recoverable strain increases and up to 4% shape recovery has been achieved. This is sufficient for many applications, although the force generated on shape recovery is only a modest 250 MPa (36,250 psi). If this can be increased to ca 700 MPa (101,500 psi), a very interesting application is possible: the use of pretrained ferrous SMA rods as prestressing tendons to produce prestressed concrete structures. In the usual fabrication of prestressed concrete, large jacks are required to stretch the steel tendons, which are then maintained in the stretched condition by massive frames situated at the ends of the molds. After the concrete has been poured and has sufficiently cured, the stretched wires are released, setting up a compressive stress in the beam or slab; it is this compressive stress which makes possible lighter, stronger, thinner concrete structures, capable of supporting much larger bending loads. This strengthening process carries the limitation of only working in a straight structure. If a pretrained ferrous SMA is employed, the stretching can be done in a factory, and there is no requirement for a restraining frame. These tendons can be shipped to the site and laid in the mold with their ends exposed. After the concrete is set, the exposed rod ends can be attached to a welding generator and heated to their recovery temperature by passing a high current through them. The recovery of the prestrain provides the prestress. The principal difference in this procedure is that these SMA prestressing tendons can, prior to being placed in the mold, be bent into any sort of curve to suit the required concrete shape, and then, when heated, provide prestressing in a curved structure. This could open up a new world of concrete structural design. Any difference in the cost of these tendons over normal steel ones would be compensated by the elimination of the heavy jacks and frames and the labor required for conventional prestressing.

The design of shape-memory devices is quite different from that of conventional alloys. These materials are nonlinear, have properties that are very temperature-dependent, including an elastic modulus that not only increases with increasing temperature, but can change by a large factor over a small temperature span. This difficulty in design has been addressed as a result of the demands made in the design of complicated smart and adaptive structures. Informative references on all aspects of SMAs are available (7–9).

BIBLIOGRAPHY

"Shape-Memory Alloys" in *ECT* 3rd ed., Vol. 20, pp. 726–736, by L. McDonald Schetky, International Copper Research Association.

1. L. C. Chang and T. A. Read, *Trans. AIME* **191**, 47 (1951).
2. W. J. Buehler, J. V. Gilfrich, and K. C. Weiley, *J. Appl. Phys.* **34**, 1567 (1963).
3. K. Otsuka, *Mater. Sci. Forum*, **56–58**, 393–404 (1990).
4. J. M. Ball and R. D. James, *Arch. Ration. Mech. Anal.* **100**, 13 (1987).
5. R. D. James, *Mater. Res. Soc. Proc.* **246**, 81–90 (1991).
6. J. D. Busch, *Proceedings of the First International Conference on Shape Memory and Superelastic Technologies*, Asilomar, Calif., 1994, p. 259.
7. C. Liang and C. A. Rogers, *J. Intell. Mater. Sys. Struct.* **1** (1990).
8. W. Tang and R. Sandström, *Int. J. Appl. Biomech.* **10(2)** (1995).
9. L. M. Schetky, in T. W. Duerig and co-workers, eds., *Engineering Aspects of Shape-Memory Alloys*, Butterworth-Heinemann, London, 1990, p. 170.
10. L. M. Wolford, D. A. Cottrell, and S. C. Karras, in Ref. 6, p. 477.
11. A. Metzger and D. Stockel, in Ref. 6, p. 401.
12. S. B. Lu and co-workers, *Shape-Memory Alloy '86: Proceedings of International Symposium of Shape-Memory Alloys*, China Academic Publishers, 1986.
13. J. P. O'Leary, J. E. Nicholson, and R. F. Gattuma, in Ref. 9, p. 477.
14. K. Yamauchi and co-workers, *Proc. Mat. Res. Soc. Jpn.* **18B**, 1159 (1993).
15. T. Duerig and co-workers, eds., *Engineering Aspects of Shape-Memory Alloys*, Butterworth-Heinemann, London, 1990.
16. H. Funukubo, ed., *Shape Memory Alloys*, Gordon and Breach Science Publishers, New York, 1987.

L. McDonald Schetky
Memry Corporation

SHORTENINGS AND OTHER FOOD FATS.

See [FATS AND FATTY OILS](#); [VEGETABLE OILS](#).

SHRIMP MEAL.

See [AQUACULTURE CHEMICALS](#); [FEEDS AND FEED ADDITIVES](#).

SIDERITE.

See [PIGMENTS](#).

SIENNA, BURNT.

See [PIGMENTS](#).

SIEVES.

See [SIZE MEASUREMENT OF PARTICLES](#).

SIGNALING SMOKES.

See [CHEMICALS IN WAR](#).

SILANES; SILANOLS.

See [SILICON COMPOUNDS](#).

SILICA

Introduction,
Amorphous silica,
Vitreous silica,
Synthetic quartz crystals,

INTRODUCTION

The term silica denotes the compound silicon dioxide [7631-86-9], SiO₂, and encompasses its various forms, including crystalline silicas, eg, quartz; microcrystalline silicas, eg, diatomaceous earth (see DIATOMITE); vitreous silicas, which are essentially supercooled liquid glasses; and amorphous silicas. The amorphous forms vary over a wide range of hydration or hydroxylation. Silica sols (colloidal silicas) are nonsettling dispersions of amorphous silica particles in a liquid, generally water. Amorphous powders can be broadly classified as either wet process or pyrogenic silicas based on the mode of manufacture. Wet process types include precipitated silicas and silica gels. Pyrogenic silicas are made at high temperature rather than from aqueous solution, and are sometimes referred to as fumed silicas.

Silicon dioxide is the most common binary compound of silicon and oxygen, the two elements of greatest terrestrial abundance. Silica constitutes about 60 wt % of the earth's crust, occurring either alone or combined with other oxides in the silicates (see SILICON COMPOUNDS, INORGANIC SILICATES). It is thus a ubiquitous chemical substance and, owing to its rich chemistry, of great geological importance. Commercially, silica is the source of elemental silicon and is used in large quantities as a constituent of building materials (see BUILDING MATERIALS, SURVEY; SILICON AND SILICON ALLOYS). In its various amorphous forms it is used as a desiccant, adsorbent, reinforcing agent, filler, and catalyst component (see DESICCANTS; FILLERS). Silica also has numerous specialized applications, eg, as piezoelectric crystals, in vitreous silica optical elements, and as glassware. Silica is a basic material of the glass (qv), ceramics (qv), and refractories (qv) industries, and an important raw material for the production of soluble silicates, silicon and its alloys, silicon carbide, silicon-based chemicals, and the silicones (see ADVANCED CERAMICS; CARBIDES; INTEGRATED CIRCUITS; SILICON COMPOUNDS, SILICONES).

Structure and Bonding

Silicon shares with the other elements of Group 14 (IVA) of the Periodic Table the property of forming an oxide of formula MO₂. This oxide shows acidic properties. Basic character becomes more pronounced for the heavier members of the Group. For example, SnO₂ is distinctly amphoteric. Like the dioxides of germanium, tin, and lead, silica is a solid of high melting point. Where silica differs is in having a three-dimensional lattice based on four-coordinate silicon. The heavier Group 14 analogues possess the more ionic structures of the rutile type. An exception is a high temperature form of GeO₂, which has four-coordinate structure. Carbon dioxide (qv), markedly different, is a gas that condenses at - 78.5° C at ca 100 kPa (1 atm) to a molecular solid having a linear OCO structure. The principal reason for the structural difference between CO₂ and SiO₂ is the ability of carbon, along with other elements of the second Period, to form strong π-bonds using valence shell *p* orbitals. Thus CO₂ assumes a molecular structure characterized by carbon–oxygen double bonds. For elements of the third and lower Periods, larger internuclear separations, ie, bond distances, make *p*_π – *p*_π overlap considerably less favorable. Because the π-bonding is accordingly diminished, the possibility of association through formation of additional σ-bonds is correspondingly enhanced. Consequently, in numerous examples of formally similar compounds, the lighter compounds are monomeric molecular species, whereas the heavier analogues of the same stoichiometry have oligomeric or polymeric structures. Among these are the analogous compounds boron trichloride, BCl₃, and aluminum chloride, Al₂Cl₃ (g); carbon disulfide, CS₂, and poly(silicon disulfide), (SiS₂)_∞; nitrogen, N₂, and phosphorus, P₄; nitric acid, HNO₃, and poly(metaphosphoric acid), (HPO₃)_∞; and nitrogen oxide (2:3), N₂O₃, and phosphorus oxide (4:6), P₄O₆. For the same reason, the silicones, (R₂SiO)_∞, form polymeric structures in preference to the monomeric structures of the analogous ketones, R₂CO.

The basic structural unit of most of the forms of silica and of the silicate minerals is a tetrahedral arrangement of four oxygen atoms surrounding a central silicon atom. Although the Si–O bond clearly possesses considerable covalent character, these materials may be treated to a good approximation according to the empirical rules dealing with the stability of ionic crystals formulated by Pauling (1), ie, the silica structure as an aggregation of Si⁴⁺ cations (ionic radius 41 pm) and O²⁻ anions (ionic radius 140 pm). The ratio of cation:anion radius (0.29) is in the range for which tetrahedral coordination of O about Si is predicted from geometric arguments based on anion–anion and cation–anion contact in coordination polyhedra (2). The SiO₂ stoichiometry requires that on the average each oxygen must be shared by silicons in two tetrahedra, whereas in accordance with the electrostatic valency rule a single oxygen cannot be shared between more than two tetrahedra. Sharing of corners is the common mode of linkage of the coordination polyhedra; sharing of edges is rarely encountered; and sharing of faces never occurs because of the decrease in stability that would result from the close approach of the silicon cations.

Structurally, silica represents a limiting case in which an infinite three-dimensional network is formed by the sharing of all oxygen atoms of a given tetrahedron with neighboring groupings. The possibility of linking tetrahedrons so that some corners remain unshared gives rise to a wide range of structural possibilities, some of which are encountered in the silicates. In structures for which all corners of the tetrahedra are not shared, each unshared oxygen atom contributes a formal negative charge to the anionic groups thus formed, which is satisfied by the presence of other cations in the silicate structure. The various ways in which the SiO₄ tetrahedra may be linked form a basis for a structural classification of the silicates and of silica itself (3). A more extensive treatment of silicate structures is available (2).

Five structural types are summarized in Figure 1. Structures containing discrete SiO⁴⁻₄ tetrahedra, mononuclear anions which are variously called silicate, orthosilicate, or tetraoxosilicate(IV) ions (4), are found in the mineral olivine, (MgFe)₂SiO₄, and zircon, ZrSiO₄. In structures containing discrete Si₂O₇ polyhedra, ie, disilicate (formerly called pyrosilicate) anions, two SiO₄ tetrahedra share a single corner. Thortveitite, Sc₂Si₂O₇, and a series of disilicates of the 4*f* (lanthanide) elements, M₂Si₂O₇, are examples of this relatively rare silicate type. Structures composed of tetrahedra sharing two oxygen atoms give rise to discrete cyclic anions such as Si₃O₉ (cyclotrisilicate), as in benitoite [1302-78-9], BaTiSi₃O₉, or Si₆O₁₂, as in beryl [1302-52-9]. Alternatively, single-chain structures may be formed as in the pyroxenes. Different spatial configurations of the single chains are possible, resulting in structures with different numbers of tetrahedra in the spatially repeating unit. Double-chain structures in which the SiO₄ tetrahedra are topologically inequivalent in the sense that some silicon atoms share two, and some three oxygen atoms, may be considered as arising from the lateral linking of single chains as in the amphiboles, in which the repeating unit is Si₄O₁₁. The composition of such double chains depends on the spatial configuration of the corresponding single chains. The sheet structures in which three oxygen atoms of each SiO₄ tetrahedron are shared, consisting of linked rings formed by six tetrahedra, is characteristic of the micas (see MICAS, NATURAL AND SYNTHETIC). A different sheet structure containing alternate four- and eight-membered rings is found in the mineral apophyllite [1306-03-2].

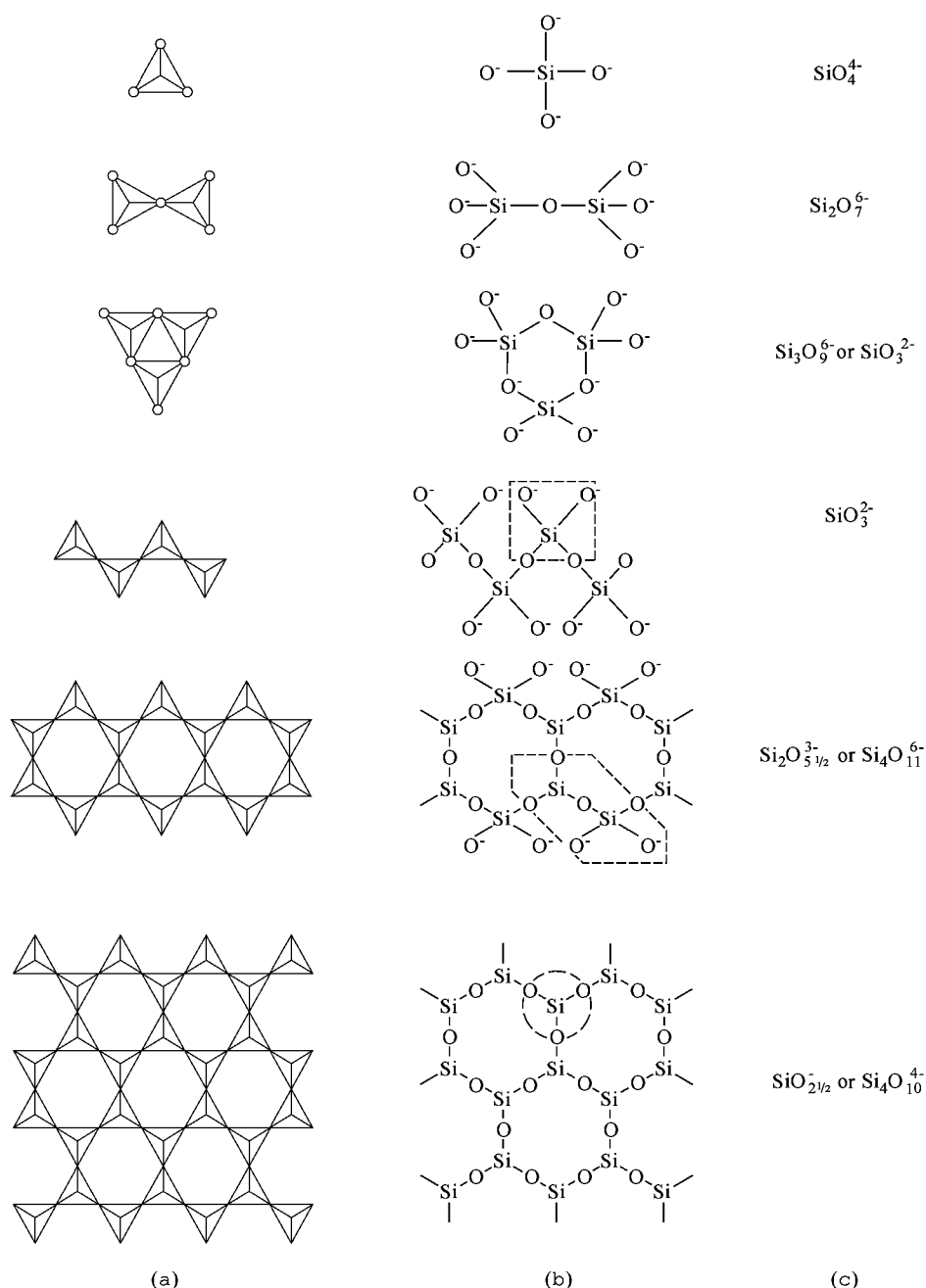


Fig. 1. Schematic representation of basic silicate structures: **(a)** modes of linkage of SiO_4 tetrahedra; **(b)** the corresponding bonding patterns; and **(c)** structural formulas (3). The Si atoms that appear to be joined to only three O atoms are joined to a fourth also, which is above the plane of the diagram.

The structures in which SiO_4 tetrahedra share all four oxygen atoms lead to the principal forms of silica. Replacement of some silicon atoms by aluminum gives a negatively charged framework of composition, $\text{Al}_x\text{Si}_{4-x}\text{O}_{20-4x}$, in which positive ions are accommodated in holes in the structure. Examples of these framework silicates include the feldspars, zeolites and ultramarines (see CLAYS; MOLECULAR SIEVES).

Although the treatment of silica and the silicates in terms of Pauling's rules suggests an ionic model, the silicon–oxygen bond in fact possesses appreciable covalency. Pauling estimated an Si–O bond distance of 163 pm (observed value 162 pm) from the sum of the covalent radii modified by the Schomaker-Stevenson correction for the electronegativity difference between Si and O, and by a shortening correction for double bonding, $d_{\pi} = p_{\pi}$, between oxygen and silicon (1,5). On the basis of this model, the residual charge on the silicon atom in SiO_4^{4-} arising from partial ionic character is reduced by π -bonding to a value in accord with the electroneutrality principle. Pauling's estimate for this residual charge is approximately +1 electronic units. From the viewpoint of a totally covalent bonding model, the silicon atom forms bonds to oxygen utilizing four equivalent sp^3 hybrid orbitals with possible participation of $p_{\pi} - d_{\pi}$ bonding. The resulting tetrahedral geometry, which follows from the assumption of sp^3 hybridization, is that found universally in the uncomplexed molecular compounds of silicon.

Because the bonding in the silicates is not dissimilar to that in other silicon compounds, it is not surprising that the various silicate structures have analogues in molecular silicon–oxygen chemistry, where oxygen bridging between silicon atoms is commonly encountered. Thus, the silicon–oxygen ring of the cyclosilicate anions is found in the organocyclosiloxanes, whereas the single-chain silicates are formally similar to linear silicone polymers, in which the unshared oxygen atoms are replaced by univalent organic groups. The organosiloxanes include, however, some structural types that have no silicate analogues, such as discrete, three-dimensional polyhedra based on the sharing of three oxygens per silicon atom (6). The relationship between the organopolysiloxanes and the silicates may be invoked to relate some structural and chemical properties of silicates to those of their organic analogues (7).

Properties

Reactions. At ordinary temperatures silica is chemically resistant to many common reagents. However, it undergoes a wide variety of chemical transformations under conditions such as high temperatures or when volatile products escape from the reaction. Reactivity is strongly dependent on the form, pretreatment, and state of subdivision of the particular sample investigated. Finely divided amorphous silica is in many circumstances considerably more reactive than bulk crystalline silica. The reactivity of high surface area amorphous silica is conditioned by the presence of surface hydroxyl (silanol) groups, half of which may be retained even after heating to 400°C (8). A large amount of information on the reactivity of these surface groups has been developed in the course of surface-chemistry investigations (9,10). Silanol groups can be expected to be present on the surface of any silica sample and to exert an influence as reactive sites. Surveys of the chemical properties are available (11,12).

Common aqueous acids do not attack silica, except for hydrofluoric acid, which forms fluorosilicate anions, eg, SiF_2^- . The rate at which the various

forms of silica are dissolved by aqueous HF decreases with increasing density, ρ, in the following sequence: vitreous silica ($\rho = 2.2 \text{ g cm}^{-3}$) < tridymite [15468-32-3] ($\rho = 2.22 \text{ g cm}^{-3}$) cristobalite [14464-46-1] ($\rho = 2.33 \text{ g cm}^{-3}$) < quartz [14808-60-7] ($\rho = 2.65 \text{ g cm}^{-3}$). Coesite [13778-38-6] ($\rho = 3.01 \text{ g cm}^{-3}$) is practically insoluble in aqueous HF. Stishovite [13778-37-5] ($\rho = 4.35 \text{ g cm}^{-3}$) is even less soluble, perhaps owing to the presence of a filled octahedral coordination shell about the silicon atom. Low quartz is attacked on the plane perpendicular to the optic axis at a rate about a hundred times greater than that on the prism faces, possibly reflecting the different surface structures on different crystal planes (13). Phosphoric acid attacks vitreous silica at elevated temperatures, forming a crystalline silicophosphate. The solubility of silica is greater in dilute than in more concentrated aqueous phosphoric acid (14). Quartz and vitreous silica are affected only slightly by aqueous alkali at room temperature. The attack is faster at higher temperatures. Precipitated amorphous silica is more reactive than vitreous silica, which in turn is more reactive than quartz.

Silica is reduced to silicon at 1300–1400°C by hydrogen, carbon, and a variety of metallic elements. Gaseous silicon monoxide is also formed. At pressures of >40 MPa (400 atm), in the presence of aluminum and aluminum halides, silica can be converted to silane in high yields by reaction with hydrogen (15). Silicon itself is not hydrogenated under these conditions. The formation of silicon by reduction of silica with carbon is important in the technical preparation of the element and its alloys and in the preparation of silicon carbide in the electric furnace. Reduction with lithium and sodium occurs at 200–250°C, with the formation of metal oxide and silicate. At 800–900°C, silica is reduced by calcium, magnesium, and aluminum. Other metals reported to reduce silica to the element include manganese, iron, niobium, uranium, lanthanum, cerium, and neodymium (16).

Of the halogens, only fluorine attacks silica readily, forming SiF₄ and O₂. A number of halogen compounds of the nonmetals and metalloids react more or less readily with silica, forming volatile silicon halogen compounds (Table 1). The formation of SiCl₄ by direct chlorination of mixtures of silica and carbon is of some technical importance.

Table 1. Reactions of Halogen Compounds and Silica

Halogen compound	Products	Conditions	Reference
HF	SiF ₄		
FNO	SiF ₄ , N ₂ O ₃	slowly at 150°C	17
SeOF ₂	SiF ₄ , SeO ₂	quantitatively	18
BrF ₃	SiF ₄ , O ₂ , Br ₂	quantitatively	19
BF ₃	SiF ₄ , (BOF) ₃	thermally	20
	SiF ₄ , cyclic (SiOF ₂) _n , (BOF) ₃ , B ₂ OF ₄ , F ₂ BOSiF ₃	microwave discharge	
CF ₃ CF ₃	SiF ₄ , CO, CO ₂	at 800°C	21
BCl ₃ , S ₂ Cl ₂ , PCl ₃	SiCl ₄	elevated temperatures	22

The acidic character of silica is shown by its reaction with a large number of basic oxides to form silicates. The phase relations of numerous oxide systems involving silica have been summarized (23). Reactions of silica at elevated temperatures with alkali and alkaline-earth carbonates result in the displacement of the more volatile acid, CO₂, and the formation of the corresponding silicates. Similar reactions occur with a number of nitrates and sulfates. Silica at high temperature in the presence of sulfides gives thiosilicates or silicon disulfide, SiS₂.

The reactions of silica with organic and organometallic compounds result in compounds containing Si–C and Si–O–C bonds. Treatment of silica with alkyl or aryl Grignard reagents, followed by hydrolysis, gives organocyclosiloxanes in high yields (24). Small amounts of low molecular weight silicon compounds can be obtained by reaction of methanol or sodium methoxide and silica over a period of days or weeks (25). Other studies indicate that fracture of silica, eg, in grinding, produces active sites that react with alcohols to form surface esters and with olefins to yield oligomerized species bonded to the surface through Si–O bonds that cannot be hydrolyzed (26).

Solubility. An important aspect of silica chemistry concerns the silica–water system. The interaction of the various forms of silica with water has geological significance and is applied in steam-power engineering where the volatilization of silica and its deposition on turbine blades may occur (see POWER GENERATION), in the production of synthetic quartz crystals by hydrothermal processes (qv), and in the preparation of commercially important soluble silicates, colloidal silica, and silica gel.

Reliable determination of the solubility of silica in water has been complicated by the effects of impurities and of surface layers that may affect attainment of equilibrium. The solubility behavior of silica has been discussed (9,27). Reported values for the solubility of quartz, as SiO₂, at room temperature are in the range 6–11 ppm. Typical values for massive amorphous silica at room temperature are around 70 ppm; for other amorphous silicas, 100–130 ppm. Solubility increases with temperature, approaching a maximum at about 200°C. Solubility appears to be at a minimum at about pH 7 and increases markedly above pH 9 (9).

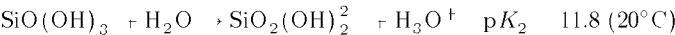
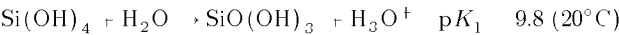
Results obtained at high temperatures indicate that the solubilities of the crystalline modifications of silica are in the order tridymite > cristobalite > quartz, an order that parallels to some extent the chemical reactivity of these forms. Lower values for solubility of crystalline as compared to amorphous silica are consistent with the free-energy differences between them.

Silica dissolves in water at high temperatures and pressures. For amorphous silica up to 200°C, the solubility in liquid water is given as follows (28):

$$c = 0.382(13.6 + t) \times 10^{-3}$$

where *c* is the concentration of dissolved silica in wt % and *t* is the temperature in °C. The solubility of quartz under pressure in H₂O (l) passes through a maximum at about 330°C, at which temperature the saturated solution contains 0.07 wt % silica, and declines rapidly as the temperature approaches the critical point (29). Vaporization of silica in steam is of potential importance to the degradation of refractories (qv), eg, in coal gasifier environments (30). The volatility of silica in steam (qv) has been discussed (31). The solubility of quartz in steam may be as high as 61.8 wt % at 963 MPa (9500 atm) and 1050°C (32–35). At high pressures and temperatures, dimeric, trimeric, and higher oligomeric forms of SiO₂ become significant molecular constituents of steam, and thus probable reactants in the SiO₂–H₂O system. Vapor species, including Si(OH)₄, Si₂O(OH)₂, and (SiO(OH)₂)₃, appear to be involved in these processes. The oligomers are more important at higher steam densities (31).

The solution in equilibrium with amorphous silica at ordinary temperatures contains monomeric monosilicic acid, Si(OH)₄. The acid is dibasic, dissociating in two steps (36):



The possibility of six-coordinate silicon species in aqueous solution has been suggested. Raman studies have indicated, however, that monosilicic acid in solution contains a tetracoordinate silicon species (37).

Solutions of monosilicic acid may also be obtained by careful hydrolysis of tetrahalo-, tetraalkoxy-, or tetraacyloxysilanes; by electrolysis or acidification of alkali silicate solutions; or by ion exchange (qv). By operating under carefully controlled conditions at low temperature and pH, solutions may be obtained that remain supersaturated with respect to amorphous silica for hours at temperatures near 0°C. Eventually, however, polymerization reactions involving the formation of siloxane linkages occur, leading ultimately to the formation of colloidal particles and further aggregation or gel

formation.

Polymerization. The basic features of silica polymerization and its technical relevance have been summarized (9). A general polymerization scheme is shown in Figure 2. When a solution of $\text{Si}(\text{OH})_4$ is formed, eg, by acidification of a solution of a soluble silicate, at a concentration greater than the solubility of amorphous silica (100–200 ppm), the monomer polymerizes to form dimers and higher molecular weight species. The mechanism is ionic, and the rate is proportional to OH^- concentration above pH 2 and to H^+ concentration below pH 2. Several mechanisms may be involved. The reaction is reported to be third order below pH 2 and second order at higher pH. The polymerization occurs in such a way as to maximize formation of siloxane linkages, Si-O-Si , forming particles with internal siloxane linkages and external SiOH groups. Above pH 7, stabilized particles (sols) grow to radii of about 100 nm. At lower pH or if salts are present to neutralize the charge on the growing particles, aggregation of particles occurs with the formation of chains and, ultimately, three-dimensional gel networks. Control of these processes by adjustment of pH and addition of coagulants is the basis of the technology for the formation of the amorphous silicas.

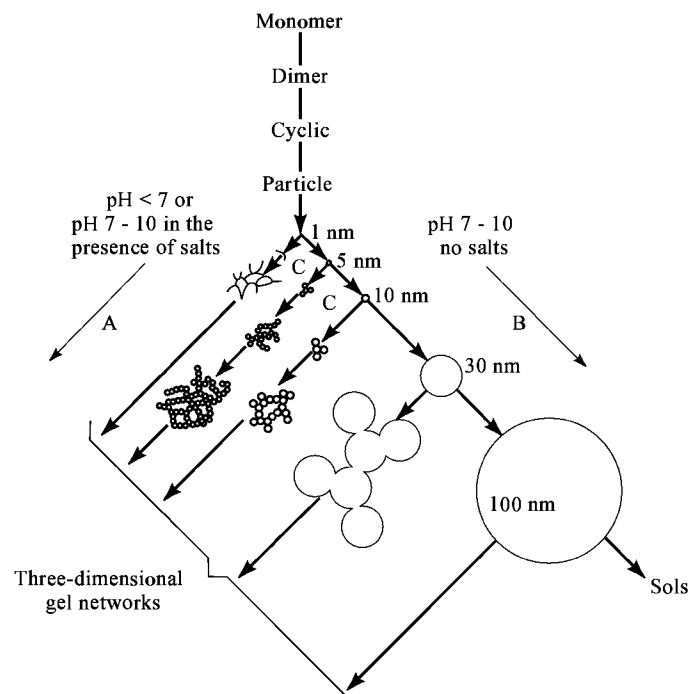
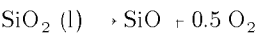


Fig. 2. Schematic representation of the polymerization of monosilicic acid (9).

Vaporization. Silica vaporizes principally by dissociation. Gaseous SiO and O_2 are the predominant vapor species. There is also some atomic oxygen and gaseous SiO_2 (38,39). The total vapor pressure over the liquid at the melting point is in the range of 1–10 Pa (10^{-4} – 10^{-5} atm). The boiling point of silica is estimated as $2797 \pm 75^\circ\text{C}$ (40). Experimental values ranging from 2230–3500°C have been reported. The heat of vaporization of SiO_2 at the melting point is given as 560 kJ/mol (134 kcal/mol), whereas the heat of the following reaction is 750 kJ/mol (179 kcal/mol):



Forms of Silica

Crystalline Silica. Silica exists in a variety of polymorphic crystalline forms (23,41–43), in amorphous modifications, and as a liquid. The literature on crystalline modifications is to some degree controversial. According to the conventional view of the polymorphism of silica, there are three main forms at atmospheric pressure: quartz, stable below about 870°C; tridymite, stable from about 870–1470°C; and cristobalite, stable from about 1470°C to the melting point at about 1723°C. In all of these forms, the structures are based on SiO_4 tetrahedra linked in such a way that every oxygen atom is shared between two silicon atoms. The structures, however, are quite different in detail. In addition, there are other forms of silica that are not stable at atmospheric pressure, including that of stishovite, in which the coordination number of silicon is six rather than four.

At the temperature limits of their stability ranges, the main forms of silica interconvert. The transformations involve a change in the secondary (nonnearest-neighbor) coordination and require the breaking and reformation of Si-O bonds. The transformation processes, known as reconstructive polymorphic transformations (44), are slow, as shown by the fact that the high temperature polymorphs can persist outside their normal stability range. The transformations are aided by or may require the presence of impurities or added mineralizers such as alkali metal oxides. Indeed, it has been suggested that tridymite cannot be formed at all in the absence of impurities, and some texts assert that pure SiO_2 occurs in only two forms: quartz and cristobalite (45).

In addition to the reconstructive transformations, each of the main forms of silica undergoes one or more transformations of a different sort, the so-called high–low, displacive, or martensitic transformations. These involve relatively small structural rearrangements, such as minor rotations of the tetrahedra without bond-breaking, which in general are facile and reversible. For example, low or α -quartz, the form stable at room temperature, transforms displacively to high or β -quartz at 574°C. This transformation is important in ceramic technology because it involves a significant volume change that can lead to cracking of ceramic bodies containing large amounts of silica. Cristobalite undergoes a similar transition at 270°C. The displacive transitions of tridymite are covered extensively in the literature (41,46,47).

The use of α and β to denote the forms of quartz has not always been uniform in the literature, particularly in older references. Although some authors have used β to denote the low form, usage as of the 1990s has the stable low temperature phase as α and the high temperature phase as β .

The melting point of high cristobalite is $1723 \pm 5^\circ\text{C}$ as based on the International Temperature Scale of 1948 (48). Because of the slowness of the tridymite–cristobalite conversion, it is possible to observe the melting point of the metastable form at 1680°C. Similarly, quartz melts at a temperature lower than that of either cristobalite or tridymite, probably about 1470°C, but the rate of fusion is comparable to the rate at which cristobalite is formed (49). The transformations among the principal crystalline forms of silica at atmospheric pressure may thus be represented in the simplified form shown in Figure 3. Corresponding thermodynamic properties of these transitions are given in Table 2; thermodynamic properties of silica forms are given in Table 3.

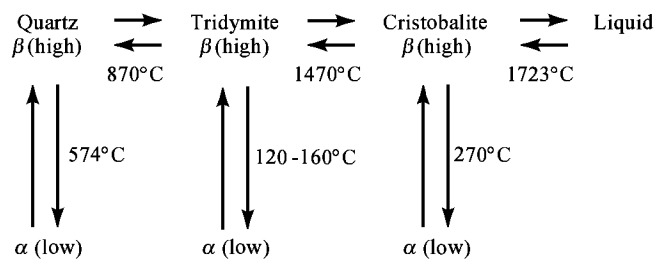


Fig. 3. The main crystallographic forms of silica stable at atmospheric pressure. The vertical directions represent the facile, displacive polymorphic transitions, whereas the horizontal directions represent the sluggish reconstructive transitions (44).

Table 2. Heat of Transformation of Various Forms of SiO₂^{a, b}

Transformation	Temperature, K	ΔH , kJ mol ^c
low quartz to high quartz	847 ± 1.5	0.73 ± 0.2
high quartz to high cristobalite	1079 ± 250	2.01 ± 0.6
low cristobalite to high cristobalite	5.3 ± 3	1.34 ± 0.3
high quartz to liquid	1696 ± 50	7.7 ± 0.8
high cristobalite to liquid	1996 ± 5	9.6 ± 2

^a Ref. 50.
^b See Fig. 3.
^c To convert J to cal, divide by 4.184.

Table 3. Thermodynamic Properties of Quartz, Cristobalite, and Liquid SiO₂^a

Property	Low quartz	Cristobalite		Liquid SiO ₂
		Low	High	
heat of formation, $\Delta H_{f, 298.15\text{ K}}$, kJ mol ^b	910.9 ± 2	908.3 ± 2	905.5	902.7
entropy, $S_{298.15\text{ K}}$, J (molK) ^b	41.5 ± 1	43.40 ± 0.13	50.05	47.93 ± 1
heat capacity, $C_{p, 298.15\text{ K}}$, J (molK) ^b	44.59	44.95	26.58	44.18

^a Ref. 50.
^b To convert J to cal, divide by 4.184.

By definition, the stable phases show no changes in properties with time at constant temperature and pressure, even when in contact with solutions or melts. Metastable or unstable phases may persist essentially indefinitely at temperatures and pressures outside their ranges of stability. The distinction is that metastable phases can be stimulated to change into a different form when placed in contact with that form; unstable phases change without the need for such contact. Thus, tridymite and cristobalite may be investigated at temperatures within the stability range of quartz and may undergo transitions of the high–low type at these temperatures. Such thermodynamically unstable but kinetically stable phases are said to be thermally stranded. An analogous situation occurs when pressure rather than temperature is the controlling variable in phase equilibria. Such phases are said to be barically stranded. For example, coesite is stable only at high pressure, but nevertheless persists when brought to atmospheric pressure and room temperature. Under particular conditions, a metastable or unstable phase may convert either to the phase that is stable under those conditions, or into some other phase that is lower in free energy than the first but also unstable or metastable. When heated to 867–1470°C, pure quartz, for example, usually converts to a disordered cristobalite rather than to tridymite.

Another representation of the stability relations of the silica minerals is shown in Figure 4. This diagram, developed in the classical studies early in the twentieth century (51), illustrates the relationship of vapor pressure to temperature. It is assumed that vapor pressure increases with temperature and that the form having the lowest vapor pressure is the most stable. The actual values of the vapor pressures are largely unknown. Therefore, the ordinate must be considered only as an indication of relative stabilities. This diagram does not show all the various forms of tridymite that have been identified.

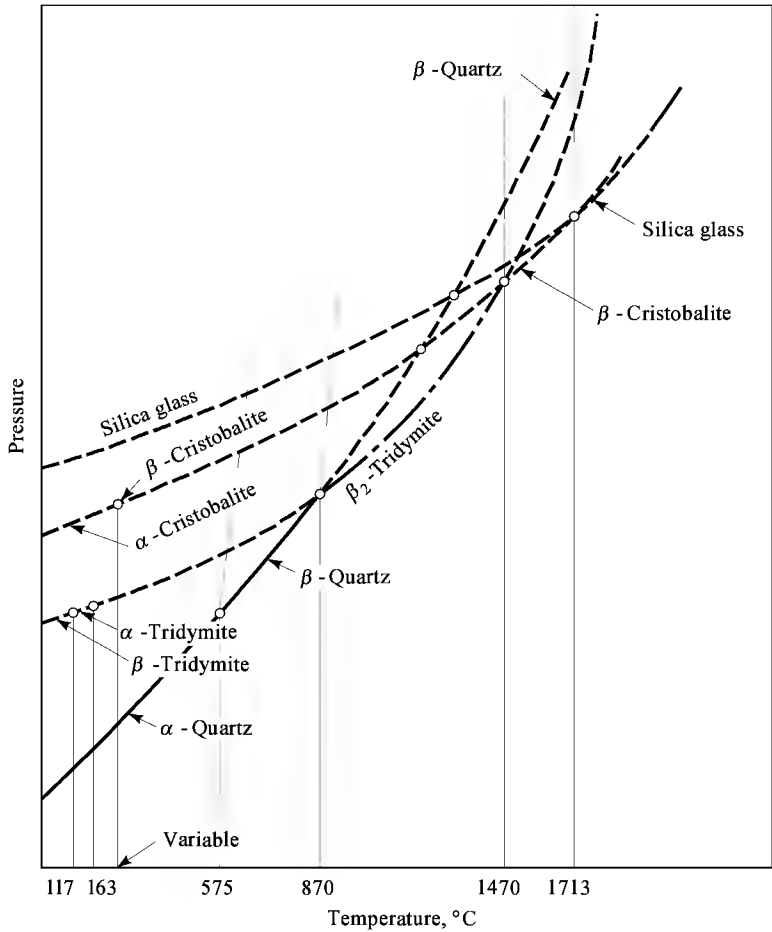


Fig. 4. Stability relations of the silica minerals.

In addition to the three principal polymorphs of silica, three high pressure phases have been prepared: keatite [17679-64-0], coesite, and stishovite. The pressure–temperature diagram in Figure 5 shows the approximate stability relationships of coesite, quartz, tridymite, and cristobalite. A number of other phases, eg, silica O, silica X, silicalite, and a cubic form derived from the mineral melanophlogite, have been identified (9), along with a structurally unique fibrous form, silica W.

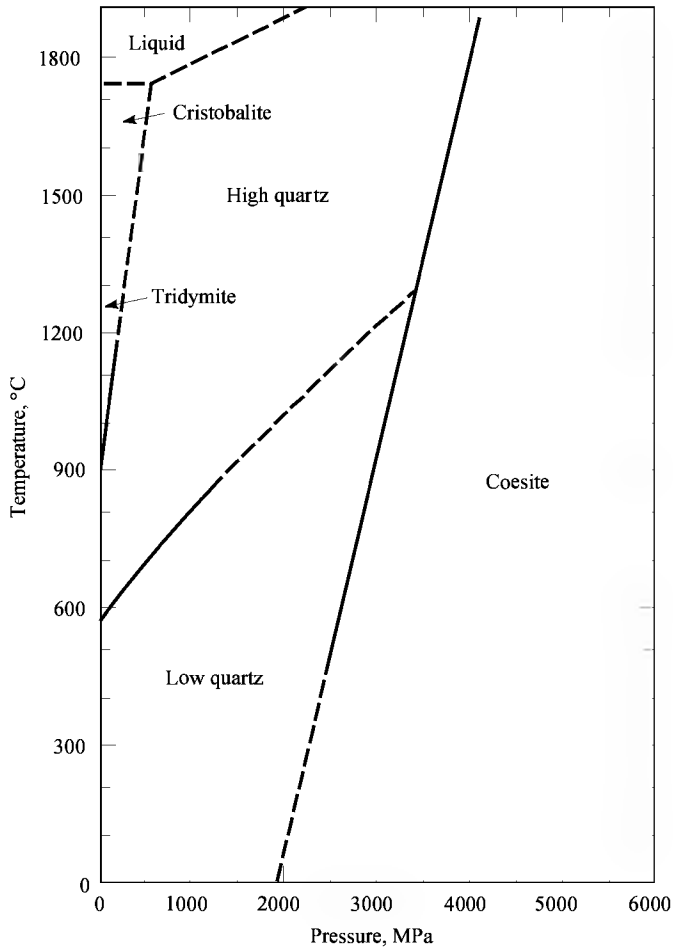


Fig. 5. Pressure–temperature diagram for the more familiar SiO₂ polymorphs (43). To convert MPa to atm, divide by 0.101.

Quartz. Low quartz (α -quartz) is common in nature as sand. Until the 1950s, natural quartz was used for commercial applications requiring

crystals. Since then, economic methods utilizing hydrothermal techniques have been developed for the synthesis of large single crystals of α -quartz (see SILICA, SYNTHETIC QUARTZ SILICA).

The atomic arrangement in high quartz (β -quartz) consists of linked tetrahedra-forming helices that, in a given crystal, are either right- or left-handed (52). The hexagonal unit cell contains three SiO_2 units having space group $P6_32$, and $a_0 = 501$ pm and $c_0 = 547$ pm at about 600°C (53). The Si–O distance is 162 pm; the density at 600°C is about 2.53 g cm^{-3} . The structure of low quartz (α -quartz) is closely similar, but somewhat less regular (Fig. 6). The unit cell has dimensions $a_0 = 491.3$ pm, and $c_0 = 540.5$ pm, space group $P3_2$, three formula units in the hexagonal unit cell, and two slightly different Si–O distances, 159.7 and 161.7 pm, in the tetrahedra (2). Density at 0°C is 2.65 g cm^{-3} . In the high–low inversion the atoms in the structure are only slightly displaced. Quartz is birefringent ($n_o = 1.5442$, $n_E = 1.5533$, Na_D line) and optically active. Individual crystals are either dextro- or levorotatory, such that $\alpha = 21.71^\circ \text{ mm}$ for the sodium D line. The sense of rotation is unchanged on passing from the low to the high form.

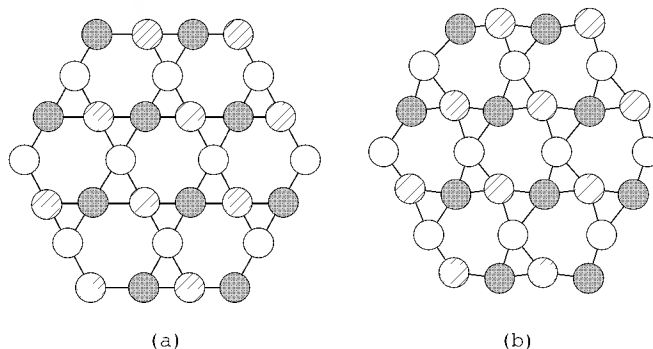


Fig. 6. Schematic illustration of structural relationships in quartz where the circles represent silicon centers only, projected on the basal plane (oxygen atoms are not shown); (○) represent the highest level, (⊗) represent the intermediate level, and (⊙) represent the lowest level. The lines are an aid to visualization only: (a) high temperature or β -form, and (b) low temperature or α -form.

The most common form of silica is low quartz, which by virtue of its piezoelectric and other properties is of considerable commercial importance. A detailed compilation of crystal structure data is available (54).

The high–low thermal inversion of quartz is accompanied by discontinuities in many physical properties, including density, indexes of refraction and birefringence, optical rotatory power, dielectric constant, and thermal expansion coefficients. It is not established, however, that all these changes are precisely coincident with the inversion point, ie, the temperature at which a crystal undergoes an abrupt, reversible change in dimensions and structure. The most common method of observing the transition is probably differential thermal analysis, by which the absorption or evolution of heat accompanying the change can be readily detected. The accepted value of the inversion temperature is $574 \pm 1.5^\circ\text{C}$ at atmospheric pressure, measured by rising temperature. The inversion temperature of natural quartz samples may vary by 40°C . A variation of as much as 160°C has been found for synthetic samples (55). This variation is believed to be associated with the formation of solid solutions with impurities. Substitution of small amounts of germanium for silicon raises the inversion temperature, whereas introduction of interstitial lithium, accompanied by substitution of aluminum to preserve charge balance, lowers it.

The pressure dependence of the high–low inversion was investigated up to 260 MPa (ca 2570 atm) (56) and up to 1 GPa (ca 9900 atm) (57). The change in the inversion temperature, ΔT , was found in the latter study to obey the equation

$$\Delta T = 1.6 + 2.871 \times 10^{-2} p - 4.284 \times 10^{-7} p^2$$

where p is the pressure expressed in 10^5 Pa (bars).

Tridymite. Tridymite is reported to be the silica form stable from 870 – 1470°C at atmospheric pressure (44). Owing to the sluggishness of the reconstructive tridymite–quartz conversion, which requires mineralizers such as sodium tungstate, alkali metal oxide, or the action of water under pressure, tridymite may persist as a metastable phase below 870°C . It occurs in volcanic rocks and stony meteorites.

Two high–low inversions in tridymite observed at 117 and 163°C (51) led to the designation of three forms, α , β_1 , and β_2 , in order of increasing temperature, as low tridymite, lower high tridymite, and upper high tridymite (46). Further work has, however, revealed a considerably more complex situation (41). Several varieties of tridymite have been suggested, including a monotropic tridymite M, which is transformed to stable tridymite S by a transition of the reconstructive type, and possibly a highly disordered tridymite U (44). Furthermore, six modifications of tridymite S are recognized, denoted S-I to S-VI in order of rising temperatures. High–low inversions occur at 64 , 117 , 163 , 210 , and 475°C . Three modifications of tridymite M also exist. Inversions are at 117 and 163°C (41).

The structure of tridymite is more open than that of quartz and is similar to that of cristobalite. The high temperature form, probably S-IV, has a hexagonal unit cell containing four SiO_2 units, where $a_0 = 503$ pm and $c_0 = 822$ pm $> 200^\circ\text{C}$, space group $P6_3 \text{ mmc}$. The Si–O distance is 152 pm. Density at 200°C is about 2.22 g cm^{-3} .

The existence of tridymite as a distinct phase of pure crystalline silica has been questioned (42,58–63). According to this view, the only true crystalline phases of pure silica at atmospheric pressure are quartz and a highly ordered three-layer cristobalite having a transition temperature variously estimated from $806 \pm 250^\circ\text{C}$ to about 1050°C (50,60). Tridymites are considered to be defect structures in which two-layer sequences predominate. The stability of tridymite as found in natural samples and in fired silica bricks has been attributed to the presence of foreign ions. This view is, however, disputed by those who cite evidence of the formation of tridymite from very pure silicon and water and of the conversion of tridymite M, but not tridymite S, to cristobalite below 1470°C (47). It has been suggested that the phase relations of silica are determined by the purity of the system (42), and that tridymite is not a true form of pure silica but rather a solid solution of mineralizer and silica (63). However, the assumption of the existence of tridymite phases is well established in the technical literature pertinent to practical work.

Cristobalite. Cristobalite is the high temperature solid form of silica, stable from 1470 – 1723°C . It is capable of metastable existence and is often obtained below 1470°C . It occurs naturally in some volcanic rocks. Pure, well-crystallized cristobalite has a high–low inversion at about 270°C , but the inversion temperature is variable and may occur as low as 170°C (64). The inversion temperature and the sharpness of the transition are dependent on the history of the sample and may be associated with the presence of foreign substances and with crystal perfection (64). The structure of high cristobalite is cubic and a_0 is 716 pm at 290°C in the eight-molecule unit cell (Fig. 7) having space group $P2_1$. The oxygen atoms of the SiO_4 tetrahedra of tridymite and cristobalite have the relationship of hexagonal close-packing to cubic close-packing, ie, the difference between the structures is similar to the difference in wurtzite and zinc blende structures. The idealized tridymite structure thus involves two-layer sequences of SiO_4 tetrahedra, and the idealized cristobalite structures, three-layer sequences (60).

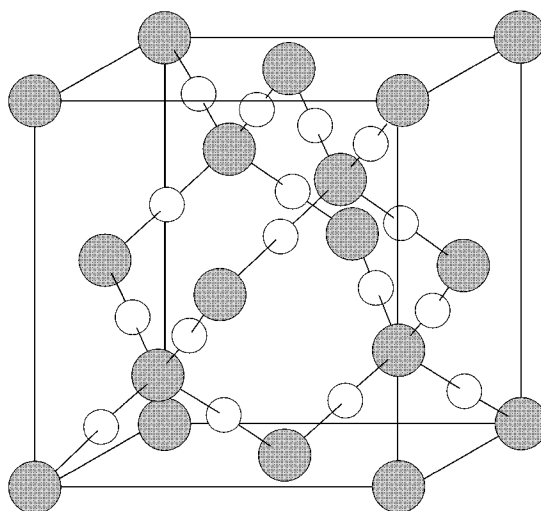


Fig. 7. Structure of high cristobalite, where $\bullet$ and $\circ$ represent Si and O, respectively (44).

Keatite. Keatite has been prepared (65) by the crystallization of amorphous precipitated silica in a hydrothermal bomb from dilute alkali hydroxide or carbonate solutions at 380–585°C and 35–120 MPa (345–1180 atm). The structure (66) is tetragonal. There are 12 SiO_2 units in the unit cell; $a_0 = 745$ pm and $c_0 = 8604$ pm; the space group is $P4_2$. Keatite has a negative volumetric expansion coefficient from 20–550°C. It is unchanged by heating at 1100°C, but is transformed completely to cristobalite in three hours at 1620°C.

Coesite. Coesite, the second most dense (3.01 g cm^{-3}) phase of silica, was first prepared in the laboratory by heating a mixture of sodium metasilicate and diammonium hydrogen phosphate or another mineralizer at 500–800°C and 1.5–3.5 GPa (14,800–34,540 atm). Coesite has also been prepared by oxidation of silicon with silver carbonate under pressure (67). The structure is monoclinic; $a_0 = b_0 = 717$ pm, $c_0 = 1.238$ pm, and $\gamma = 120^\circ$. There are 16 SiO_2 units per unit cell; the space group is $C2/c$. Coesite persists as a basically stranded phase at atmospheric pressure and is a stable phase at high pressure. Density values are $2.93\text{--}3.01 \text{ g cm}^{-3}$ (43). The existence of a true equilibrium between coesite and quartz has been demonstrated. Pressure–temperature diagrams for the coesite–quartz equilibrium have been summarized (23). Coesite has been found in nature in the meteor crater in Arizona.

Stishovite. Stishovite was first prepared (68) in the laboratory in 1961 at 1200–1400°C and pressures >16 GPa (158,000 atm). It was subsequently discovered, along with natural coesite, in the Arizona meteor crater. It has been suggested that these minerals are geological indicators of meteorite impact structures. Stishovite ($\rho = 4.35 \text{ g cm}^{-3}$) is the densest known phase of silica. The structure, space group $P4_2/mnm$, is similar to that of rutile, TiO_2 , and the silicon atom is octahedrally coordinated by six oxygens (69). Four Si–O distances in the octahedra are 176 pm; two are 181 pm. Stishovite is the least reactive of the many forms of silica.

Silica W. Silica W is the lightest silica phase ($\rho = 1.97 \text{ g cm}^{-3}$). It was prepared by the disproportionation of silicon monoxide (70). It forms microcrystalline fibers and differs radically from all other silica phases because the SiO_4 tetrahedra share edges rather than corners. The structure consists of parallel chains and is analogous to that of SiS_2 and SiSe_2 . Silica W reacts rapidly with traces of water vapor, transforming to amorphous silica with silica tetrahedra sharing corners rather than edges.

Microcrystalline Silicas. Various microcrystalline (cryptocrystalline) materials such as flint, chert, and diatomaceous earth are found in nature (see DIATOMITE). These may arise from amorphous silica, often of biogenic origin, which undergoes compaction and microcrystallization over geologic time.

Noncrystalline Silicas. The noncrystalline forms of silica include bulk vitreous silica and a variety of other amorphous types, which are of substantial commercial importance (see SILICA, AMORPHOUS SILICA; SILICA, VITREOUS SILICA). A review discussing the occurrence, synthesis, properties, and applications of the various forms of silica adsorbents is available (71).

Vitreous silica (silica glass) is essentially a supercooled liquid formed by fusion and subsequent cooling of crystalline silica. It is found in nature in fulgurites, ie, fused bodies resulting from lightning striking quartz sand.

Liquid silica is highly viscous, and supercooling to the glassy form occurs readily. In practice, vitreous silica is prepared by fusion of crystalline quartz or quartz sand. The primary source of high quality silica sand used in the United States is Brazil, although commonly available domestic quartz sand is available for applications not requiring highest purity. Vitreous silica is also made by flame or plasma hydrolysis of silicon tetrachloride, by thermal decomposition of silicate esters, or by sputtering of SiO_2 . Glasses prepared by flame fusion processes may contain significant (>1000 ppm) amounts of hydroxyl impurity, which affect optical transmission as well as thermal and mechanical properties.

The structure of vitreous silica is a continuous random network of SiO_4 tetrahedra, linked through the sharing of corners. It differs from crystalline silica in having a broader distribution of Si–O–Si bond angles and a more random distribution of one tetrahedron with respect to another (44). The density is 2.2 g cm^{-3} .

The properties of high quality vitreous silica that determine its uses include high chemical resistance, low coefficient of thermal expansion ($5.5 \times 10^{-7} \text{ }^\circ\text{C}^{-1}$), high thermal shock resistance, high electrical resistivity, and high optical transmission, especially in the ultraviolet. Bulk vitreous silica is difficult to work because of the absence of network-modifying ions present in common glass formulations. An extensive review of the properties and structure of vitreous silica is available (72).

Amorphous silica exists also in a variety of forms that are composed of small particles, possibly aggregated. Commonly encountered products include silica sols, silica gels, precipitated silica, and pyrogenic silica (9,73). These products differ in their modes of manufacture and the way in which the primary particles aggregate (Fig. 8). Amorphous silicas are characterized by small ultimate particle size and high specific surface area. Their surfaces may be substantially anhydrous or may contain silanol, $-\text{SiOH}$, groups. These silicas are frequently viewed as condensation polymers of silicic acid, $\text{Si}(\text{OH})_4$.

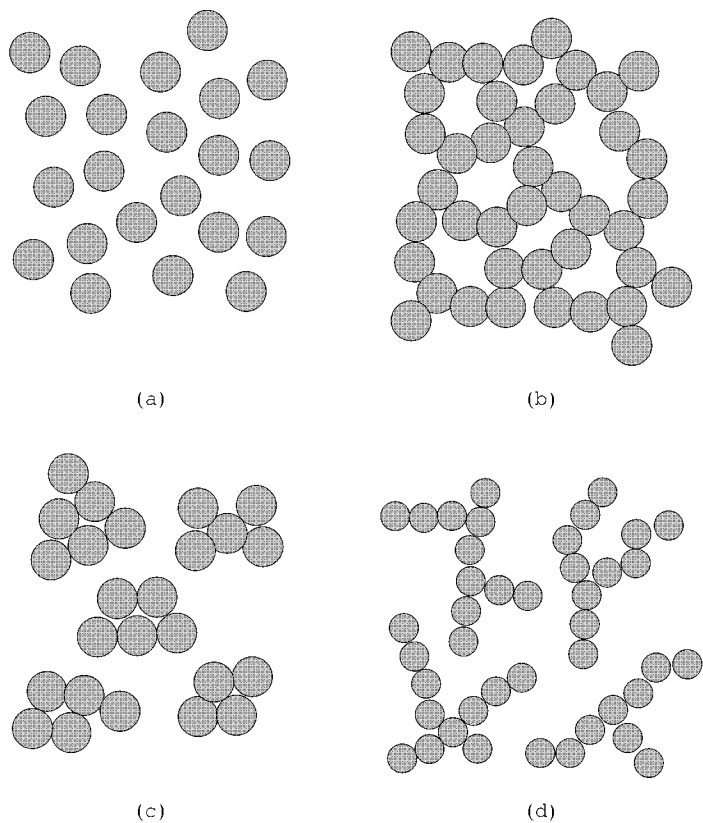


Fig. 8. Different forms of commercially important amorphous silica: (a) sol, (b) gel, (c) precipitate, and (d) pyrogenic.

Silica sols are often called colloidal silicas, although other amorphous forms also exhibit colloidal properties owing to high surface areas. Sols are stable dispersions of amorphous silica particles in a liquid, almost always water. Commercial products contain silica particles having diameters of about 3–100 nm, specific surface areas of 50–270 m² g, and silica contents of 15–50 wt %. These contain small (<1 wt%) amounts of stabilizers, most commonly sodium ions. The discrete particles are prevented from aggregating by mutually repulsive negative charges.

Silica sols are typically manufactured by passing a dilute solution of sodium silicate through a hydrogen ion exchange resin (Fig. 9). As sodium is removed, the silicate solution becomes unstable and silica begins to polymerize. Under the right conditions, ie, not too concentrated and the absence of coagulating ions, the growing silica particles do not aggregate. Particle size can be controlled in various ways, for example by accreting additional monomeric silica on existing particles, taking care to avoid excessive addition rates of reactants that would nucleate new particles rather than accrete onto existing ones. Many variations have been proposed. The final products range from faintly opalescent liquids in the lower particle size ranges to milky dispersions in the higher particle size ranges. All but the lowest particle sizes can be made at concentrations of 50% silica, which nevertheless remain stable for many months or even years and flow almost like water.

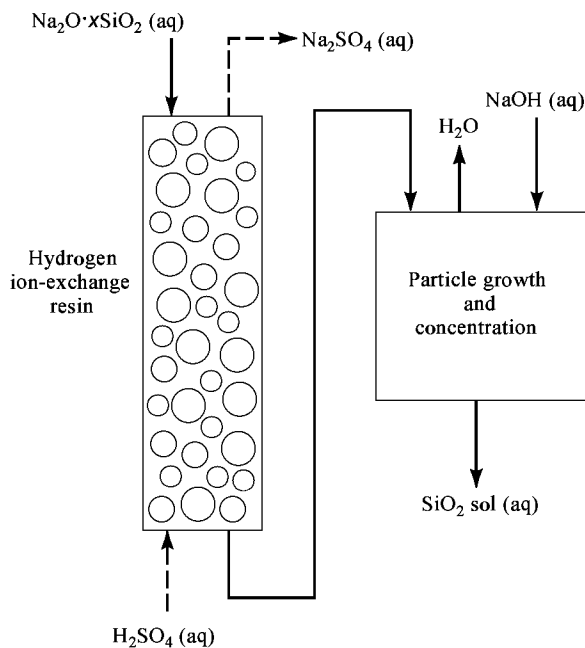


Fig. 9. Schematic of the manufacturing process for silica sols.

Silica gels contain three-dimensional networks of aggregated silica particles of colloidal dimensions. Gels are typically manufactured by acidifying a relatively concentrated aqueous solution of sodium silicate (Fig. 10). The material that first forms is called a hydrosol, and consists of growing silica polymers. The hydrosol is typically formed at low (<4) pH, and if subsequent washing steps are also conducted at relatively low pH, the final product then has a high surface area (>700 m² g). Such high surface areas are indicators of small primary particle size (1–3 nm). If the hydrosol is formed at higher pH, the primary particles grow to larger size and a lower surface area results.

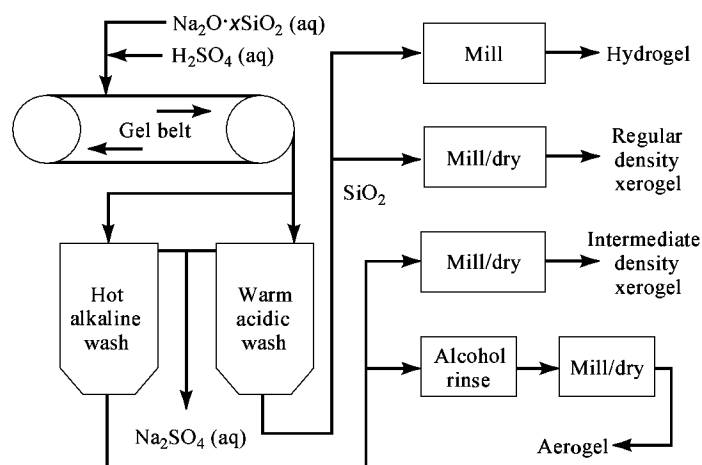


Fig. 10. Schematic of the manufacturing process for silica gels.

When the hydrosol ceases to flow like a liquid (the gel time), it is termed a hydrogel (Fig. 11a). As formed, the pores are filled with the medium (usually water) in which the gel is prepared. The hydrogel may be washed to remove the by-product salt and sold in that form, in which case it may consist of up to 70% water. Because the water is trapped in the pores, the final product can still be a relatively free-flowing powder.

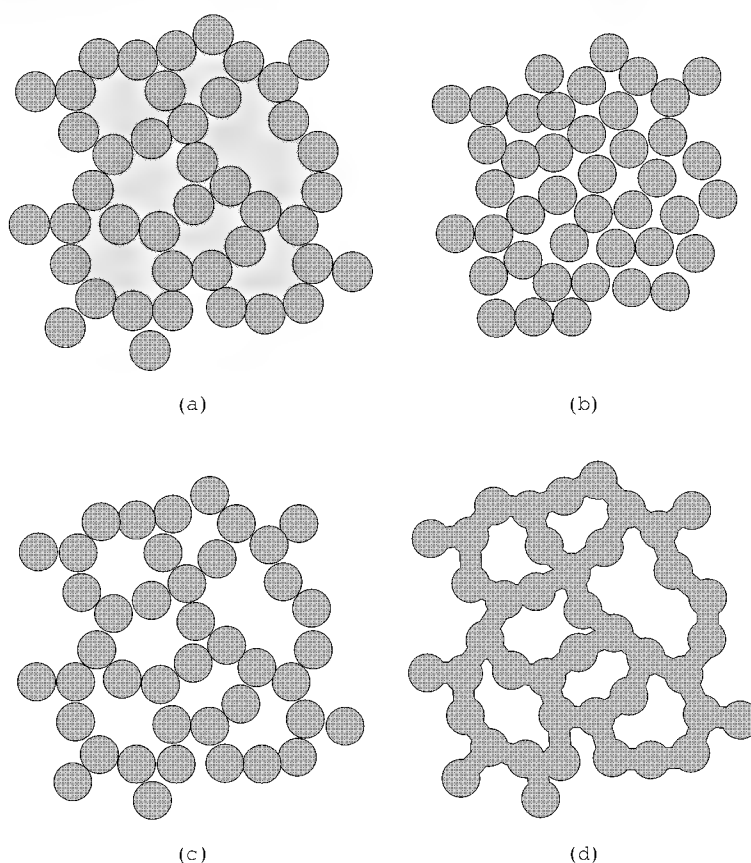


Fig. 11. Different forms of silica gel: (a) hydrogel, where the enclosed lightly shaded areas represent water; (b) regular density xerogel; (c) aerogel; and (d) intermediate density xerogel.

The hydrogel may be converted into various other forms by altering the washing and drying processes (74). Simple removal of the water results in extensive shrinkage owing to surface tension forces. Silica gels dried this way are termed regular density xerogels (Fig. 11b). The remaining pores are generally too small for the gel to function as an adsorbent for anything but small molecules; for this reason, such products are excellent desiccants.

If the water in the pores is replaced by a liquid of much lower surface tension, eg, an alcohol, which is subsequently removed by heating, the shrinkage forces are much reduced and a higher pore volume results. If such a gel is heated under pressure above the critical temperature of the liquid, resulting in the disappearance of the liquid–vapor interface, surface tension effects are absent and a very voluminous dry silica gel is obtained. Gels prepared in such a way as to minimize shrinkage are termed aerogels (qv; see SUPPLEMENT) (Fig. 11c). Very high pore volumes can be achieved, but aerogels shrink if they are wetted with water and redried. Aerogels have been reviewed (75,76).

Dry gels of relatively high pore volume can also be prepared without resorting to solvents and high pressures. The more common method, called hydrothermal treatment, is to wash the hydrogel for prolonged times in hot, sometimes alkaline water. Alternatively, the hydrosol may be formed at higher pH, ie, >4 and often considerably higher, a region where the gel time is much shorter. This is how silica gel beads are made. Two mechanisms have been proposed to explain the resultant reduced surface area and higher pore volume; there appears to be evidence to support both.

A solution redeposition mechanism has been described in detail (9). Because the solubility of silica in water increases rapidly above pH 8, reinforcement can occur through the deposition of dissolved silica at the juncture of primary particles owing to the reduction of solubility in areas having a negative radius of curvature. The effect is to cement the primary particles together enough so that they resist collapse upon drying. A silica made in this way is referred to as an intermediate density xerogel because the density of the dried silica structure is intermediate between that of the regular (high) density xerogel and the low density aerogel.

An alternative mechanism involves a physical reorientation of particles at constant total volume to form clumps that have much larger pores (74). Electron micrographs of successive changes in morphology as a function of steeping time in hot water are offered in support of this interpretation.

Silica gels may also be produced by the hydrolysis and polycondensation of silicon alkoxides, eg, tetraethylorthosilicate, $\text{Si}(\text{OC}_2\text{H}_5)_4$. This is often

referred to as a sol–gel process, and is used to produce silica that gels in a mold. High purity is possible. Applications include the manufacture of porous ceramics, membranes, and encapsulated molecules (77) (see SOL–GEL TECHNOLOGY).

Precipitated silicas are powders obtained by coagulation of silica particles from an aqueous medium under the influence of high salt concentrations or other coagulants. Under typical conditions, the primary particles grow to sizes larger than 4–5 nm and are coagulated into aggregates by the sodium ion contributed by the sodium silicate raw material. In fact, this self-coagulation must be avoided if the objective is to produce a stable sol instead of a precipitate. In contrast to the conditions in a gel, the entire liquid phase is not enclosed by the solid silica phase.

A typical manufacturing process for precipitated silicas is given in Figure 12. The precipitated slurry is washed to remove soluble salts, typically in filter presses. The conditions of washing, although important, have less effect on the final product properties than for silica gels. The filter cake is dried by one of several methods, such as spray drying or rotary drying. The dried product may also be milled or granulated to achieve specific desired agglomerate particle sizes. Many variations are possible, resulting in a wide range of surface areas, agglomerate particle sizes, and purities.

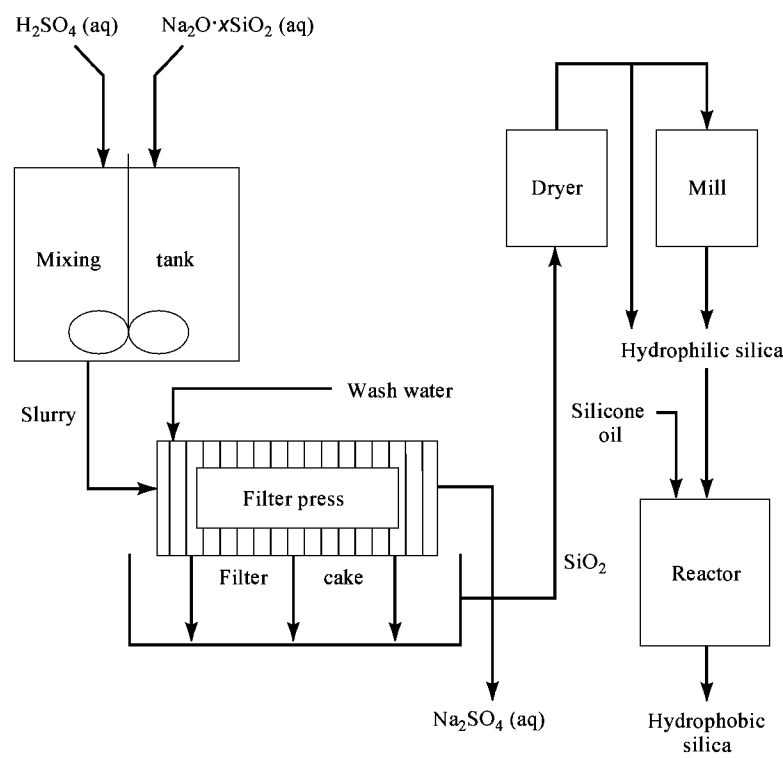
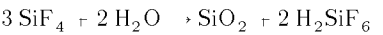


Fig. 12. Schematic of a manufacturing process for precipitated silicas.

Figure 12 shows an optional step at the end of the process in which the silica is chemically reacted with a silicone oil, typically polydimethylsiloxane, to render the product hydrophobic. Other aftertreatments are also commonly employed, such as wax coatings that enhance the performance of precipitated silicas used as flatting agents in paints (78).

An alternative route of manufacture is from silicon tetrafluoride that is generated in large quantities as a by-product of the production of phosphate fertilizers. The reaction is



Precipitated silicas having properties matching those from the conventional sodium silicate route are said to be possible (79). The hydrofluosilicic acid that is generated can be recycled in the system, or can be used to manufacture sodium silicofluoride or ammonium fluoride.

Pyrogenic silicas are often called fumed silicas, but common usage refers to them by the trade names Aerosil (Degussa) and Cab-O-Sil (Cabot). These are produced by vapor-phase processes, generally by the vapor-phase hydrolysis (ca 1000°C) of silicon tetrachloride. Other methods include vaporization of SiO₂, vaporization and oxidation of Si, and high temperature oxidation and hydrolysis of silicon compounds such as silicate esters. Although the particles are formed at temperatures where crystalline phases are stable, x-ray diffraction reveals that pyrogenic silicas are entirely amorphous.

Owing to the absence of liquid water at the time of formation, the surface of pyrogenic silica particles has fewer silanol groups per unit area than the wet process amorphous silicas. This gives rise to different agglomeration behavior in liquid dispersions than is exhibited for precipitated silicas and silica gels. For this reason, pyrogenic silicas are exceptionally good thixotropic agents in certain systems. Thickening results from the formation of a network of particles, which is enhanced if the particles have both hydrophilic and hydrophobic areas (80).

Economic Aspects

The estimated worldwide production of various forms of amorphous silica and the principal suppliers as of the early 1990s are given in Table 4. The annual U.S. production of sand is about 27 million metric tons (72).

Table 4. Worldwide Annual Production and Suppliers for Amorphous Silicas^a

Type of silica	Production, t × 10 ³	Suppliers
sols	21	Nalco, DuPont, Eka Nobel
gels	90	Grace (Davison), PQ, SCM, Unilever (Crosfield), Asahi, Akzo-PQ
precipitated	800	Degussa, J. M. Huber, PPG, Rhône-Poulenc, Akzo-PQ, Tokuyama Soda, Nippon Silica
pyrogenic	100	Degussa, Cabot, Wacker, Tokuyama Soda

^a Refs. 81–83.

Health and Safety Factors

An overview of silica in biological systems is available (9). Silica is chemically inert, and in bulk it is relatively biologically inert. It is listed in the *U.S. Food*

Chemicals Codex and *National Formulary*, and is permitted as an additive in food for human and animal consumption subject to regulatory restrictions.

Inhalation of crystalline or fused vitreous silica dust, usually over long periods, causes a disabling, progressive pulmonary disease known as silicosis (84). Amorphous silicas have not been linked to silicosis (85), but can cause respiratory irritation. The history and politics of silicosis have been reviewed (86). Standards have been set or recommended for occupational exposures (87,88) and review articles on the health effects of silica are available (83,89).

Silica has also come under scrutiny as a possible carcinogen. The International Agency for Research on Cancer (IARC) published a monograph in 1987 (90) concluding that evidence for the carcinogenicity of crystalline silica to experimental animals is sufficient; for the carcinogenicity of amorphous silica to experimental animals is inadequate; for the carcinogenicity of crystalline silica to humans is limited; and for the carcinogenicity of amorphous silica to humans is inadequate. The results of additional studies on the relationship between silica and cancer have been published (91). Diatomaceous earth, widely used as a filter aid, has been receiving more attention on health and safety issues from various U.S. governmental agencies because of the presence of microcrystalline silica.

Ingested silica can dissolve in the digestive tract and is rapidly eliminated in the urine (92). In humans, the estimated oral lethal dose of silica is $\sim 15\text{ g kg}$ (93). Powdered silica absorbs moisture and oils from the skin. It can cause chapping and skin irritation, especially in sensitive individuals. Good local and general exhaust is essential for working with powdered silicas. Respirators certified for dust protection should be available for use in instances where dust cannot be controlled with containment and or ventilation. Loose-fitting rubber gloves with cotton liners should be worn when handling silica.

Uses

The diversity of silica forms and their properties leads to a broad range of applications. Reviews on usage include those on crystalline and vitreous silicas (94), on silica sols (95), on silica gels and precipitated silicas (96), and on pyrogenic silicas, precipitated silicas, and silica gels (97).

Crystalline Silica. Quartz sand is of course the principal raw material for the production of glass (qv). Cristobalite and β -quartz are used in glass ceramics (qv), ie, ceramics produced by the controlled crystallization of glass. Silica is a main constituent of ceramics (qv). For example, refractory silica brick containing small amounts of Al_2O_3 is used as roof brick for open-hearth furnaces at temperatures $\sim 1600^\circ\text{C}$ (see REFRACTORIES). Silica sand or flour (ground quartz) is the raw material for soluble silicates, such as sodium silicate, which is consistently ranked as one of the top 50 U.S. industrial chemicals (98) (see SILICON COMPOUNDS, SYNTHETIC INORGANIC SILICATES).

Because of its piezoelectric properties, synthetic α -quartz is used for frequency control in electrical oscillators and filters and in electromechanical transducers. When mechanically stressed in the correct direction, α -quartz develops an electric polarization. The opposite is also true: an applied electric field gives rise to a mechanical distortion in the crystal. Thin sections of quartz are cut to dimensions that produce the desired resonance frequency when subjected to an alternating electric field; the vibrating crystal then reacts with the driving circuit to produce an oscillation that can be narrowly controlled. Quartz is ideal for this application because it is hard, durable, readily synthesized, and can be tuned to high accuracy, for example, quartz crystal clocks can be made that are stable to one part in 10^9 .

Vitreous Silica. Silica is the basic raw material of the glass industry. The vitreous silica structure forms the basis of commercial glass compositions, whose properties are modified by the addition of other metal oxides (see GLASS). Because of its chemical and thermal resistance, vitreous silica is used in laboratory glassware (up to 1000°C), in furnaces and radiant heaters, and as lamp envelopes. Silica fibers are used in precision instruments, eg, balances and thermal expansion apparatus. Thin films (qv) of vitreous silica are applied to dielectric components of integrated surfaces. Because of optical transparency, particularly in the ultraviolet region, vitreous silica is used for prisms, cells, windows, and other optical components. An application of increasing importance is the use of vitreous silica in optical waveguides used for long distance telecommunications. Its low thermal expansion has made vitreous silica a material of choice for astronomical telescope mirror blanks. In space technology, fused silica has been used in the windows on every U.S. peopled spacecraft (86), as well as in the thermal protection tiles on the space shuttle orbiters (99).

Amorphous Silica. Silica sols are used as binders and stiffeners, for modifying the frictional properties of fibers, eg, as an antislip agent on shipping boxes, and to improve the processing characteristics of textile fibers; as reinforcing polymers, eg, in the investment casting process used to produce molds for metal objects such as golf clubs; as viscosity agents; as a component of coatings, eg, for plastics, photographic films, and glass lenses; in inorganic–organic compositions to impart strength, eg, added to isophorone diisocyanate [4098-71-9] and a pentaerythritol triacetate–tetraacetate mixture to produce a material for a polycarbonate plate (100); and as polishing agents. An interesting application of such fine particles as polishing agents is in the preparation of silicon wafers for semiconductor applications. Extreme smoothness is necessary to enable the circuits that are put on the wafer to have the smallest possible dimensions. The mechanism of polishing is generally believed to be chemical–mechanical; the high pH of the sol leads to oxidation of the silicon surface, followed by mechanical removal of the oxidized layer under the action of the silica particles and the polishing pad. A side benefit is that the abrasive is chemically similar to the material being polished, reducing the chance for contamination that might affect the electrical properties of the final chip. Thus, silica sols are an important contributor to the computer revolution.

Silica gels are used to modify adhesives and the viscosity and thixotropy of liquids, as adsorbents and catalyst supports, and for other related purposes (9). The commercial use of gels as adsorbents is rather extensive. Regular density xerogels serve as desiccants. A more recent application of primary commercial importance is the adsorption of haze-forming proteins from beer (qv). This process is called chillproofing because the haze forms when the beer is cooled after a period of aging subsequent to bottling. The haze is caused by a delayed reaction between naturally occurring proteins and polyphenols that originate in the malt. Silica hydrogels (101) and intermediate density xerogels (102) are employed because both have pores of sufficient size ($\sim 10\text{ nm}$ average diameter) to admit the haze-forming proteins. The adsorption is selective; foam-stabilizing proteins, considered desirable by brewers, are excluded.

Hydrogels are used in the refining of edible oils to adsorb phospholipids, trace metals, and soaps (103). The adsorption capacity depends on the ease of hydration of the adsorbates, so best performance demands careful control of moisture content in the system (104). Silica hydrogel in combination with alumina has been found to be useful for purifying used cooking oils in order to extend their life and enhance the quality of the fried food (105).

Silica gels are also used as selective adsorbents in column chromatography (qv) (106). An exceptionally narrow particle size distribution is necessary to prevent excessive backpressure and to produce sharp peaks in the chromatogram. Precipitated silicas are not suitable because their agglomerates break down much more easily than do the stronger particles found in gels. High purity is also important to prevent contamination of the product being separated. The pore size distribution must be matched to the separation being attempted. The surface of the silica often reacts with a silane to create a reverse-phase packing, so-called because the elution sequence of solutes is the reverse of that for a normal-phase packing, ie, an untreated silica.

Silica gels are used in the paint (qv) and coatings industry as flattening agents. The silica particles protrude from the surface, and the resulting increase in roughness reduces the gloss of the coating. Particle size and degree of dispersion are obviously of importance. Wax coatings are sometimes added. These improve the redispersibility of silica gel particles that settle in the can during storage and improve the scratch resistance of the dry coating (78). Pyrogenic silicas enjoy wide use as thixotropic agents in coatings, inks, and resins.

Fine silica particles are incorporated in the surface of plastic films to prevent the adhesion of two sheets in contact with one another. The application is called antiblocking. The relatively low refractive index of the silica particles makes them more difficult to detect.

Both silica gels and precipitated silicas are used in toothpaste as abrasives and thickeners. Silicas are particularly useful in the production of clear gel-type toothpastes because it is possible to match the index of refraction of the liquid components of the formulation to that of the silica to achieve transparency. This cannot be done with conventional abrasive agents, such as calcium carbonate. Because both gels and precipitates can be manufactured having high pore volumes, these are also used as carriers for flavors, feed supplements such as choline chloride, pesticides (qv), etc. Gels, precipitates, and pyrogenic silicas are all used as anticaking and free-flow agents in products such as fire extinguisher powders, cocoa, milk substitutes, spices, animal feeds, etc.

Precipitated silicas are used extensively as reinforcing fillers. The largest use as of the mid-1990s is in rubber used in the production of tires. This has become a rapidly growing market ever since the development of tire formulations that rely more heavily on silicas to offer lower rolling resistance, and thus higher gasoline mileage. Carbon black has been the traditional reinforcing agent in tires. Silica can be used to replace or complement the carbon. Moreover, the exclusive use of silica permits the production of durable rubber goods in colors other than black, eg, shoe soles. The degree of reinforcement increases

with the quantity of bound rubber, ie, rubber that cannot be extracted with solvents, which forms by a free-radical mechanism during milling. Pyrogenic silicas have also been used for rubber reinforcement, but the higher cost has generally limited these silicas to higher value formulations, eg, silicone rubbers.

Precipitated and pyrogenic silicas are widely employed in antifoams used in industry. To be effective, the silica surface must be rendered hydrophobic by reaction with an agent such as polydimethylsiloxane. The hydrophobic particles are dispersed in a carrier fluid such as mineral or silicone oil. The mechanism of bubble breaking is disputed (107,108). One theory is that the mechanism is dewetting of the silica particle by the foam lamella, which creates a defect in the film that leads to its rupture. The criterion for dewetting is a three-phase contact angle of 90° or more. The three phases are the aqueous foam lamella, the solid silica particle, and the carrier oil. The properties of precipitated silica that optimize its performance in the largest U.S. use, ie, pulp (qv) and paper (qv) defoaming, have been identified for the two common methods of hydrophobing: dry-roast (109) and *in situ* (110). Other defoaming applications include paint and coatings, textile dye baths, and, particularly in Europe, laundry detergents.

Two more recent applications for amorphous silicas are expected to grow to large volumes. Precipitated silicas are used in the manufacture of separator sheets placed between cells in automotive batteries. Their function is to provide a controlled path for the migration of conductive ions as a result of the porosity of the silica particles. Additionally, both precipitated silicas and aerogels are being developed for use in low temperature insulation, where the low thermal conductivity of the dry silica powders makes them useful in consumer products such as refrigerators (83).

BIBLIOGRAPHY

"Silica and Silicate Minerals," treated under "Silica and Inorganic Silicates," in *ECT* 1st ed., Vol. 12, pp. 268–303, by G. W. Morey, Geophysical Laboratory, Carnegie Institution of Washington; "Silica (Introduction)" in *ECT* 2nd ed., Vol. 18, pp. 46–61, by T. D. Coyle, National Bureau of Standards; in *ECT* 3rd ed., Vol. 20, pp. 748–766, by T. D. Coyle, National Bureau of Standards.

1. L. Pauling, *The Nature of the Chemical Bond*, 3rd ed., Cornell University Press, Ithaca, N.Y., 1960.
2. A. F. Wells, *Crystal Chemistry*, 4th ed., Oxford University Press, London, 1975, Chapt. 23.
3. W. E. Addison, *Structural Principles in Inorganic Compounds*, John Wiley & Sons, Inc., N.Y., 1963, p. 141.
4. International Union of Pure and Applied Chemistry (IUPAC), *Nomenclature of Inorganic Chemistry*, 2nd ed., Butterworths, London, 1970.
5. L. Pauling, *J. Phys. Chem.* **56**, 361 (1952).
6. A. J. Barry and co-workers, *J. Am. Chem. Soc.* **77**, 4248 (1955).
7. W. Nell, *Angew. Chem. Intern. Ed. Engl.* **2**, 73 (1963).
8. G. J. Young, *J. Colloid Sci.* **13**, 67 (1958).
9. R. K. Iler, *The Chemistry of Silica*, John Wiley & Sons, Inc., New York, 1979.
10. H. P. Boehm, *Adv. Catal.* **16**, 179 (1966).
11. R. Calas, P. Pascal, and J. Wyart, *Nouveau Traite de Chimie Minerale*, Vol. 8, Pt. 2, Masson et Cie, Paris, France, 1965.
12. *Gmelins Handbuch der Anorganischen Chemie*, Vol. 15, Pt. B, Verlag Chemie GmbH, Weinheim, Germany, 1959.
13. F. M. Emsberger, *J. Phys. Chem. Solids*, **13**, 347 (1960).
14. V. N. Sveshnikova and E. P. Damlova, *Zh. Neorg. Khim.* **2**, 928 (1957).
15. H. L. Jackson, F. D. Marsh, and E. L. Muetterties, *Inorg. Chem.* **2**, 43 (1963).
16. F. Trombe and M. FOLAX, *Compt. Rend.* **216**, 268 (1943).
17. O. Ruff, W. Menzel, and W. Neumann, *Z. Anorg. Chem.* **208**, 293 (1932).
18. E. B. R. Prideaux and C. B. Cox, *J. Chem. Soc.* 739 (1928).
19. H. J. Emeleus and A. A. Woolf, *J. Chem. Soc.* 164 (1950).
20. P. Baumgarten and W. Bruns, *Chem. Ber.* **74B**, 1232 (1941).
21. F. E. Brinckman and G. Gordon, *Proceedings of the International Symposium on Decomposition of Organometallic Compounds to Refractory Ceramics, Metals, and Metal Alloys*, Dayton, Ohio, 1967.
22. L. White and O. K. Rice, *J. Am. Chem. Soc.* **69**, 267 (1947).
23. E. M. Levin, C. R. Robbins, and H. F. McMurdie, *Phase Diagrams for Ceramists*, American Ceramic Society, Columbus, Ohio, 1964, and supplements 1969, 1975, and 1981.
24. **Ger. Pat. 1,028,784** (Apr. 24, 1958), H. Kautsky.
25. E. Daubaeh, *Z. Naturforsch.* **8B**, 58 (1953).
26. R. E. Benson and J. E. Castle, *J. Phys. Chem.* **62**, 840 (1958).
27. W. Stober, *Adv. Chem. Ser.* **67**, 161 (1967).
28. A. S. Berezhnoi, *Silicon and Its Binary Systems*, Consultants Bureau, New York, 1960, p. 137.
29. G. C. Kennedy, *Econ. Geol.* **45**, 639 (1950).
30. M.-C. Cheng and I. B. Cutler, *J. Am. Ceram. Soc.* **62**, 593 (1979).
31. J. W. Hastie, *High Temperature Vapors*, Academic Press, Inc., New York, 1975.
32. G. C. Kennedy and co-workers, *Am. J. Sci.* **260**, 501 (1962).
33. A. I. Semenova and D. S. Tsiklis, *Zh. Fiz. Khim.* **44**, 205 (1970).
34. E. L. Brady, *J. Phys. Chem.* **57**, 706 (1953).
35. O. Glemser and H. C. Wendtlandt, *Adv. Inorg. Chem. Radiochem.* **5**, 215 (1963).
36. S. A. Greenberg, *J. Chem. Ed.* **36**, 218 (1959).
37. D. Fortnum and J. O. Edwards, *J. Inorg. Nucl. Chem.* **2**, 264 (1956).
38. R. F. Porter, W. A. Chupka, and M. C. Inghram, *J. Chem. Phys.* **23**, 216 (1955).
39. L. Brewer and D. F. Mastick, *J. Chem. Phys.* **19**, 834 (1951).
40. H. L. Schick, *Chem. Rev.* **60**, 331 (1960).
41. R. B. Sosman, *The Phases of Silica*, Rutgers University Press, New Brunswick, N.J., 1965.
42. N. A. Toropov and co-workers, *Handbook of Phase Diagrams of Silicate Systems*, Vol. II, Israel Program for Scientific Translations, Jerusalem, Israel, 1972.
43. C. Frondel, *Dana's System of Mineralogy*, Vol. 3, John Wiley & Sons, Inc., New York, 1962.
44. F. A. Cotton and G. Wilkinson, *Advanced Inorganic Chemistry*, 5th ed., Wiley-Interscience, New York, 1988.
45. W. D. Kingery, H. K. Bowen, and D. R. Uhlmann, *Introduction to Ceramics*, 2nd ed., John Wiley & Sons, Inc., N.Y., 1976.
46. R. B. Sosman, *The Properties of Silica*, American Chemical Society Monograph 37, Reinhold Publishing Corp., New York, 1927.
47. V. G. Hill and R. Roy, *Trans. Brit. Ceram. Soc.* **57**, 496 (1958).
48. S. J. Schneider, *Compilation of the Melting Points of the Metal Oxides*, National Bureau of Standards Monograph 68, Washington, D.C., 1963.
49. Ref. 41, Chapt. 7.
50. Joint Army–Navy–Air Force (JANAF), *Thermochemical Tables*, 2nd ed., NSRDS– NBS 37, 1971.
51. C. N. Fenner, *Am. J. Sci.* **36**, 331 (1913).
52. W. G. Moffatt, G. W. Pearsall, and J. Wulff, *The Structure and Properties of Materials*, Vol. 1, John Wiley & Sons, Inc., New York, 1964.
53. R. W. G. Wyckoff, *Crystal Structures*, Vol. 1, 2nd ed., Wiley-Interscience, New York, 1963.
54. "Crystal Data, Determinative Tables," in J. D. H. Donnay and H. M. Ondik, eds., *Inorganic Compounds*, Vol. 2, 1973; H. M. Ondik and A. D.

Mighell, eds., *Inorganic Compounds*, Vol. 4, National Bureau of Standards Joint Committee on Powder Diffraction, Washington, D.C., 1978.

55. M. L. Keith and O. F. Tuttle, *Am. J. Sci.* **250A**, 203 (1952).

56. R. E. Gibson, *J. Phys. Chem.* **32**, 1197 (1928).

57. H. S. Yoder, *Trans. Am. Geophys. Union*, **31**, 827 (1950).

58. Ref. 28, p. 117.

59. O. W. Flurke, *Silikattechn.* **12**, 304 (1961).

60. O. W. Flurke, *Ber. Deut. Keram. Ges.* **32**, 369 (1955).

61. Y. E. Budnikov and Y. E. Pivinskii, *Russ. Chem. Rev.* 210 (1967).

62. W. F. Ford, *The Effect of Heat on Ceramics*, MacLaren & Sons, London, 1967, pp. 82–83.

63. R. Wollast, *Proceedings of the 8th Conference on the Silicate Industry*, Akad  miai Kiado, Budapest, Hungary, 1966.

64. O. W. Flurke, *Ber. Deut. Keram. Ges.* **33**, 319 (1956).

65. O. W. Flurke, *Ber. Deut. Keram. Ges.* **33**, 319 (1956).

66. J. Shropshire, P. P. Keat, and P. A. Vaughan, *Z. Kryst.* **112**, 409 (1959).

67. L. Coes, *Science*, **118**, 131 (1953).

68. S. M. Stishov and S. V. Popova, *Geokhimiya*, 837 (1961).

69. S. M. Stishov and N. V. Belov, *Dokl. Akad. Nauk*, **143**, 951 (1962).

70. A. Weiss and W. Weiss, *Z. Anorg. Allgem. Chem.* **276**, 95 (1954).

71. S. Kondo, *Studies Surf. Sci. Catal.* **99**, 93–113 (1996).

72. R. Br  ckner, *J. Non-Cryst. Solids*, **5**, 123, 177 (1970).

73. G. Alexander, *Silica and Me*, American Chemical Society, Washington, D.C., 1967.

74. D. Barby, in G. D. Parfitt and K. S. W. Sing, eds., *Characterization of Powder Surfaces*, Academic Press, Inc., New York, 1976, pp. 353–425.

75. J. Fricke and A. Emmerling, *Adv. Mater.* **3**(10), 504 (1991).

76. R. Mehrotra, *J. Indian Chem. Soc.* **70**, 301–303 (1993).

77. L. C. Klein and R. H. Woodman, *Key Engr. Mtk.* **115**, 109–124 (1996).

78. D. Aldcroft and G. J. Earl, *Paintindia*, **45**(5), 19–28 (1995).

79. P. D. Agrawal, *Chem. Eng. World*, **30**(1), 31–33 (1995).

80. Ref. 9, p. 589.

81. P. Kleinschmit, in R. Thompson, ed., *Ind. Inorg. Chem.: Prod. Uses*, Royal Society of Chemistry, Cambridge, U.K., 1995, Chapt. 12, pp. 327–349.

82. Technical data, R. Patterson, The PQ Corp.

83. R. Glenn, *Ceram. Eng. Sci. Proc.* **13**(3–4), 153–159 (1992).

84. R. J. Lewis, *Sax's Dangerous Properties of Industrial Materials*, 8th ed., Van Nostrand Reinhold Co., Inc., New York, 1992.

85. *Silicosis—Caused by Amorphous Silica?* Technical Bulletin Pigments No. 76, Degussa Corp., Frankfurt, Germany, 1988.

86. D. Rosner and G. Markowitz, *Deadly Dust*, Princeton University Press, Princeton, N.J., 1991.

87. M. Sittig, *Handbook of Toxic and Hazardous Materials*, Noyes Publications, Park Ridge, N.J., 1981.

88. *Recommended Health-Based Limits in Occupational Exposure to Selected Mineral Dusts (Silica, Coal)*, World Health Organization Technical Report Series 734, Geneva, Switzerland, 1986.

89. D. S. Feigin, *J. Thorac. Imag.* **4**(1), 68–80 (1989).

90. *Silica and Some Silicates*, IARC Monographs Vol. 42, Oxford University Press, Oxford, U.K., 1987.

91. *Occupational Exposure to Silica and Cancer Risk*, IARC Monographs Vol. 97, Oxford University Press, Oxford, U.K., 1990.

92. F. Sauer and co-workers, *Can. J. Biochem. Phys.* **37**, 183–191 (1959).

93. Joint FAO WHO Expert Committee on Food Additives, *WHO Food Add. Ser.* (5), 21–30 (1974).

94. G. H. Beall, *Rev. Mineral.* **29**, 469–505 (1994).

95. C. C. Payne, in H. E. Bergna, ed., *The Colloid Chemistry of Silica*, American Chemical Society, Washington, D.C., 1994, pp. 581–594.

96. R. E. Patterson, in Ref. 93, pp. 617–626.

97. H. K. Ferch, in Ref. 93, pp. 481–504.

98. *Chem. Eng. News*, **74**(26), 41 (June 24, 1996).

99. L. J. Korb and co-workers, *Ceram. Bull.* **60**, 1188 (1981).

100. **Rus. Pat. 60/137, 939** (1985) (to Asahi Glass Co., Ltd.).

101. **U.S. Pat. 3,617,301** (Nov. 2, 1971), D. Barby and J. P. Quinn.

102. **U.S. Pat. 5,149,533** (Sept. 22, 1992), K. A. Berg, R. H. Witt, and R. Derolf.

103. **U.S. Pat. 4,629,588** (Dec. 16, 1987), W. A. Welsh and Y. O. Parent.

104. A. Nock, *Proceedings of the 86th American Oil Chemists' Society meeting*, San Antonio, Tex., 1995.

105. **U.S. Pat. 5,391,385** (Feb. 21, 1995), J. C. Seybold.

106. K. K. Unger, *Porous Silica: Its Properties and Use as Support in Column Liquid Chromatography*, Elsevier Science Publishing Co., Inc., New York, 1979.

107. R. J. Pugh, *Adv. Coll. Interface Sci.* **64**, 67–142 (1996).

108. R. D. Kulkarni, E. D. Goddard, and P. Chandar, in Prud'homme, ed., *Foams*, Surfactant Science Series 57, Marcel Dekker, Inc., New York, 1996, pp. 555–585.

109. R. E. Patterson, in *1988 Nonwoven Conference*, TAPPI Press, Atlanta, Ga., 1988, pp. 39–48.

110. R. E. Patterson, *Colloids Surf. A*, **74**, 115–126 (1993).

General References

R. K. Iler, *The Chemistry of Silica*, John Wiley & Sons, Inc., New York, 1979.

R. B. Sosman, *The Phases of Silica*, Rutgers University Press, New Brunswick, N.J., 1965.

Robert E. Patterson
The PQ Corporation

AMORPHOUS SILICA

Amorphous silica, ie, silicon dioxide [7631-86-9], SiO₂, does not have a crystalline structure as defined by x-ray diffraction measurements. Amorphous silica, which can be naturally occurring or synthetic, can be either surface-hydrated or anhydrous. Synthetic amorphous silica can be broadly divided into two categories of stable materials (1): vitreous silica or glass (qv), which is made by fusing quartz at temperatures greater than approximately 1700°C (see SILICA, VITREOUS SILICA), and microamorphous silica, which is discussed herein.

Microamorphous silica includes silica sols, gels, powders, and porous glasses. These consist of ultimate particles of the inorganic polymer (SiO₂)_n, where a silicon atom is covalently bonded in a tetrahedral arrangement to four oxygen atoms. Each of the four oxygen atoms is covalently bonded to at least one silicon atom to form either a siloxane, -Si-O-Si-, or a silanol, -Si-O-H-, functionality. The bond distances and bond angles in amorphous silica are similar to those of cristobalite [14464-46-1] (2): Si-O bond distances are approximately 0.16 nm, and Si-O-Si bond angles are about 148°. Surface silanol groups can be isolated from one another, so that intramolecular hydrogen bonding does not occur (Fig. 1a); vicinal to one another, thus promoting the formation of intramolecular hydrogen bonding (Fig. 1b); or geminal to one another, whereby two silanol groups are bonded to the same silicon atom (Fig. 1c). Initially formed low molecular weight species condense to form ring structures so as to maximize siloxane and minimize silanol bonds (Fig. 2).

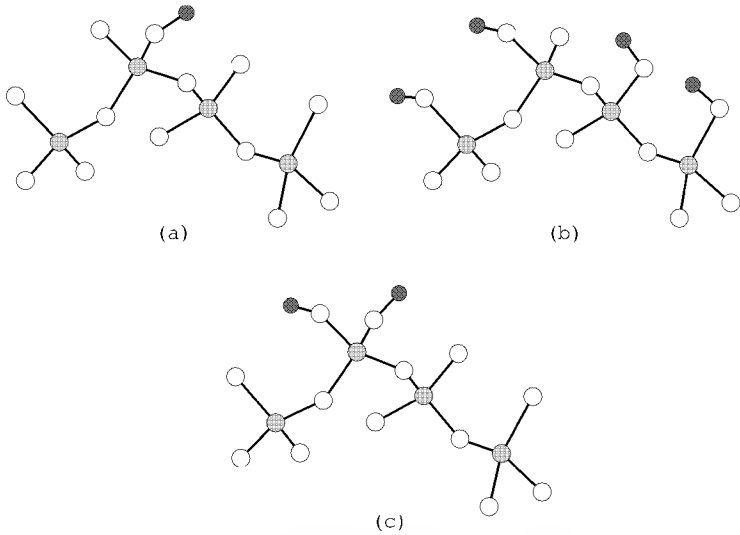


Fig. 1. Silanol groups of amorphous silica surface, where Si = Si; O = O; and H = H: (a) isolated, (b) vicinal, and (c) geminal.

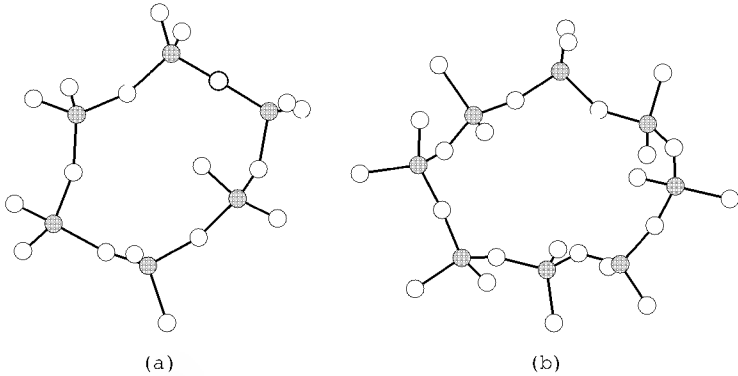


Fig. 2. Silica ring structures, where Si = Si; O = O: (a) 12-membered SiO₂ hexamer, and (b) 16-membered SiO₂ octamer.

A random arrangement of rings leads to the formation of complex structures of generally spherical particles less than approximately 100 nm in diameter (3) (Fig. 3). These particles have high surface area values, generally greater than about 3 m² g. Crystallinity is usually determined spectroscopically by using x-ray diffraction (4) to establish the absence or presence of definitive lines in the diffraction pattern of silicas. For example, Figure 4 shows x-ray diffraction patterns of crystalline quartz sand (Fig. 4a) and of cristobalite (Fig. 4b). Both exhibit distinct reflections that can be used for identification. The x-ray diffraction patterns of four commercial synthetic amorphous silicas are shown in Figure 5. The colloidal silica (Fig. 5a) is a sol or stable dispersion of discrete particles. There is a general absence of definitive reflections in these curves. Studies of radial distribution function curves of a vitreous silica have indicated that regions of local atomic order can exist which approximate the cristobalite structure. Whereas amorphous silica is closely related to the cristobalite structure, this local order is believed to be limited to crystalline domains of up to 2 nm in diameter that have completely random orientations and which are thus statistically distributed (5). Sharp x-ray diffraction patterns (lines) would therefore not be obtained.

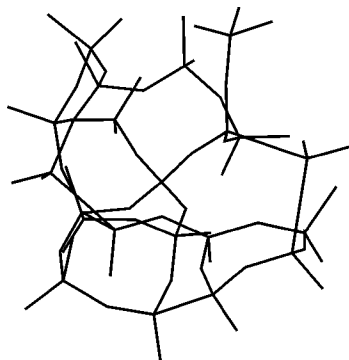


Fig. 3. Complex ring structures in SiO₂ polymer.

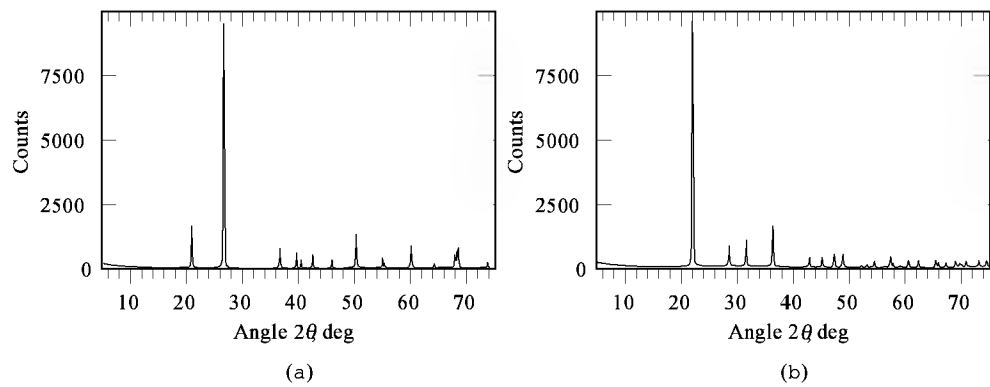


Fig. 4. X-ray diffraction patterns using a Cu anode: (a) crystalline quartz sand, and (b) cristobalite.

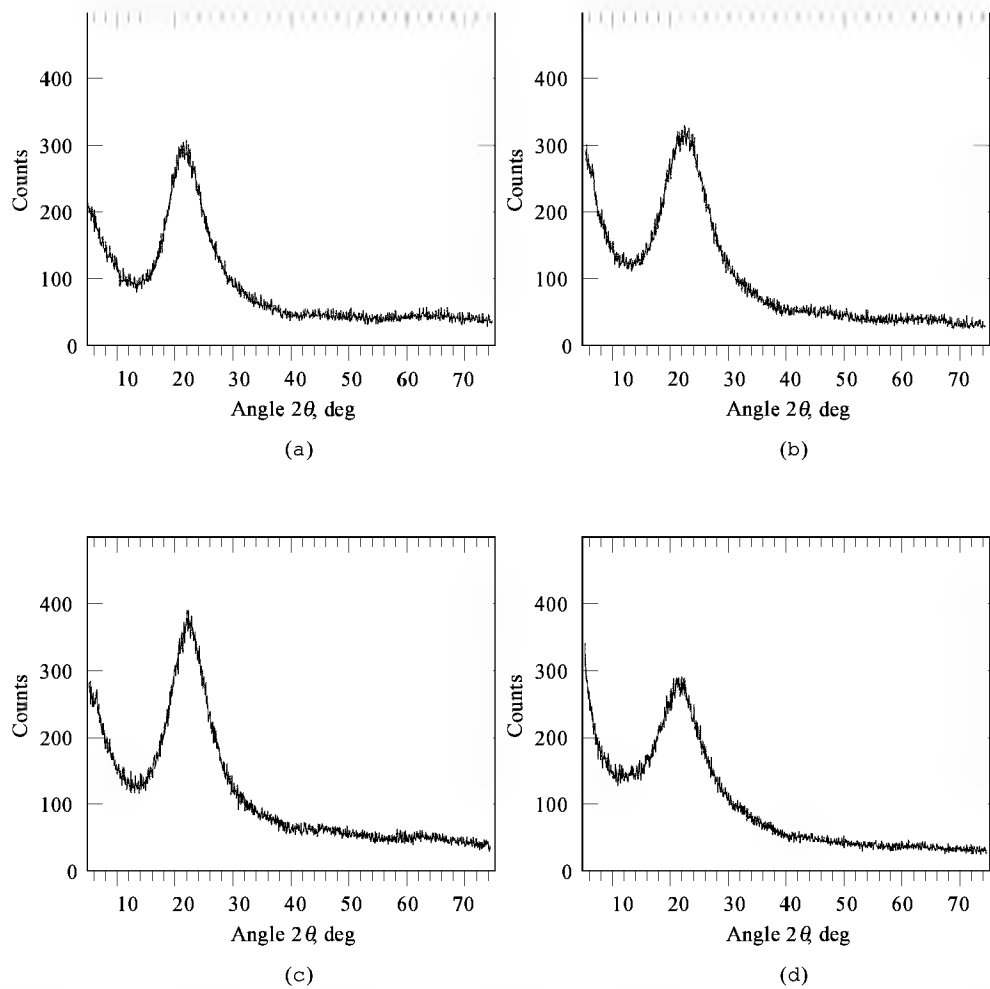


Fig. 5. X-ray diffraction pattern of commercial samples: (a) colloidal silica, DuPont Ludox TM50; (b) silica gel, Sylodent 700; (c) precipitated silica, PPG Hi-Sil 190; and (d) pyrogenic (fumed) silica, Wacker HDK.

Chemical methods to determine the crystalline content in silica have been reviewed (6). These are based on the solubility of amorphous silica in a variety of solvents, acids or bases, with respect to relatively inert crystalline silica, and include differences in reactivity in high temperature fusions with strong bases. These methods are qualitative, however, and fail to satisfy regulatory requirements to determine crystallinity at 0.1% concentration in bulk materials.

Microamorphous silica can be divided into microparticulate silica, ie, microscopic sheets and fibers, and highly hydrated silica (1). The

microparticulate silicas are the most important group commercially. These include silicas precipitated from aqueous solution and silicas formed in the vapor phase, called pyrogenic or fumed silica. Several synthetic routes exist to prepare any form of microporous silica, where microporous often refers to the ability of the silica to adsorb gases or liquids. For example, pyrogenic silica can be prepared by either (1) vaporizing silicon dioxide in an arc or in a plasma jet, and condensing it in an inert gas; (2) oxidizing the more volatile silicon monoxide, SiO , in the vapor phase with air, and condensing it; or (3) oxidizing silicon compounds (qv), such as SiH_4 , SiCl_4 , or HSiCl_3 , in the vapor phase in a hydrocarbon flame, and then condensing to produce branched-chain aggregates of silica up to approximately $1\text{ }\mu\text{m}$ in length (Fig. 6). Pyrogenic silica is collected as a finely divided powder.

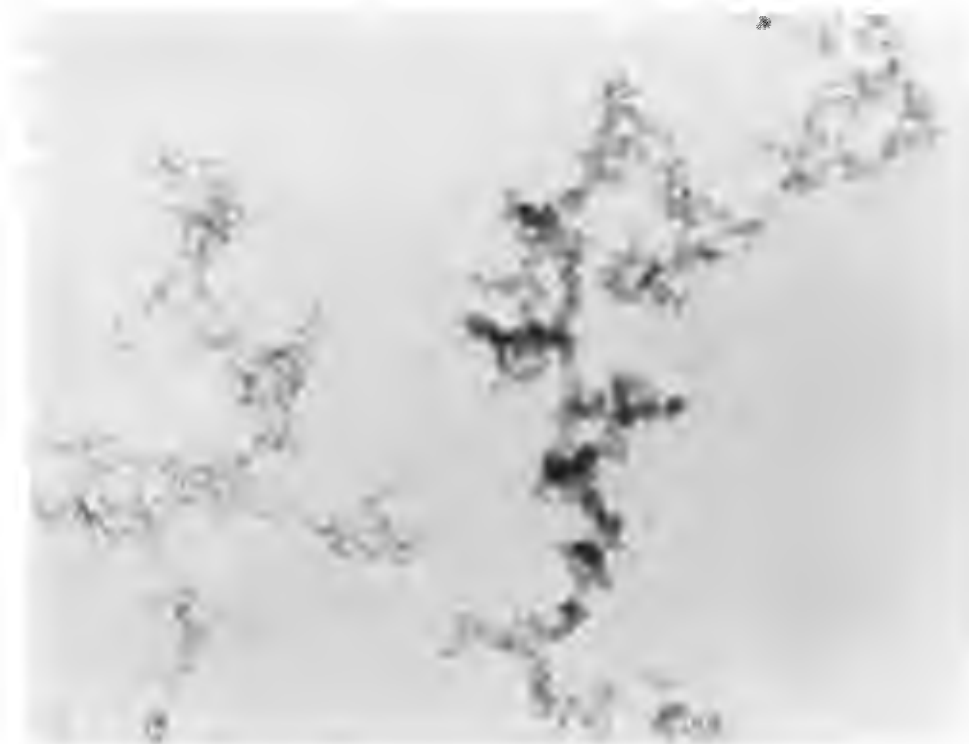


Fig. 6. Transmission electron micrograph of a commercial pyrogenic silica, Wacker HDK. Magnification of $225,000\times$.

Amorphous silica formed in aqueous solution can occur as sols, gels, or particles. A silica sol can either consist of polysilicic acid having a molecular weight of SiO_2 up to about 100,000, or be a stable dispersion of fine colloidal particles of diameters $>5\text{ nm}$ (Fig. 7). A silica gel has a three-dimensional, continuous structure (Fig. 8). Macroscopic particles formed by aggregation of a sol or very small particles of a gel to afford bunched aggregates of silica up to approximately $1\text{ }\mu\text{m}$ in diameter are called precipitated silica (Fig. 9). This is collected as a powder by physically agglomerating these aggregates and evaporating the water.

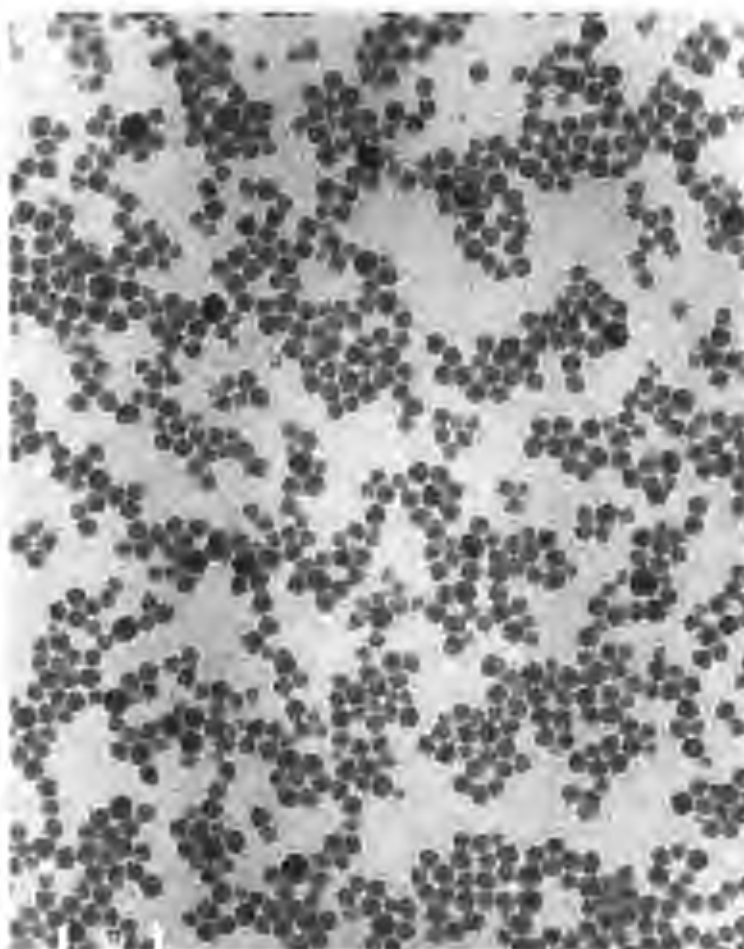


Fig. 7. Transmission electron micrograph of a commercial colloidal silica, DuPont Ludox TM50. Magnification of $225,000\times$.



Fig. 8. Transmission electron micrograph of a commercial silica gel. Magnification of 300,000 $\times$.

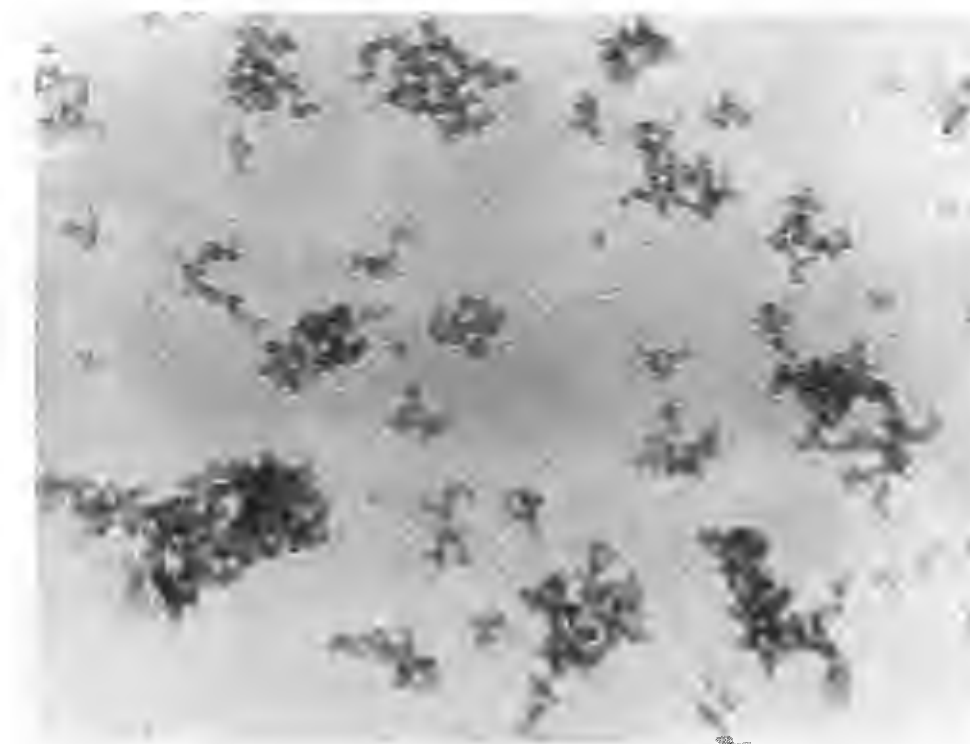


Fig. 9. Transmission electron micrograph of a commercial precipitated silica, PPG Hi-Sil 190. Magnification of 225,000 $\times$.

Microscopic sheets of amorphous silica have been prepared in the laboratory by either (1) hydrolysis of gaseous SiCl_4 or SiF_4 to form monosilicic acid [10193-36-9] (orthosilicic acid), $\text{Si}(\text{OH})_4$, with simultaneous polymerization in water of the monosilicic acid that is formed (7); (2) freezing of colloidal silica or polysilicic acid (8–10); (3) hydrolysis of HSiCl_3 in ether, followed by solvent evaporation (11); or (4) coagulation of silica in the presence of cationic surfactants (12). Amorphous silica fibers are prepared by drying thin films of sols or oxidizing silicon monoxide (13). Hydrated amorphous silica differs in solubility from anhydrous or surface-hydrated amorphous silica forms (1) in that the former is generally stable up to 60°C, and water is not lost by evaporation at room temperature. Hydrated silica gel can be prepared by reaction of hydrated sodium silicate crystals and anhydrous acid, followed by polymerization of the monosilicic acid that is formed into a dense state (14). This process can result in a water content of approximately one molecule of H_2O for each silanol group present.

Characterization

Amorphous silica is distinguished by chemical composition, physical properties, and characteristics of the particles. Exact analytical procedures depend on the type and intended application of the amorphous silica. The most important chemical information is the amount of silica, % SiO_2 ; percentage of associated water, determined by weight loss at 105°C; total solids content of nonoxidizable materials, determined by weight loss upon ignition at approximately 1200°C; presence of stabilizers, such as formaldehyde (1); carbon content, such as carbon dioxide, carbonate, and organic carbon determined by combustion to carbon dioxide; level of soluble salts, such as chloride and sulfate, determined by x-ray fluorescence spectroscopy (15) or by chemical methods such as extraction; level of nonsiliceous ash; level of metal impurities, determined by x-ray fluorescence, eg, aluminum, sodium, calcium, manganese, and iron, or by atomic absorption, eg, copper and chromium; and silanol group density, determined by infrared spectroscopy (16) or chemical

titration.

Physical characteristics include measurements of pH determined as a 5% aqueous slurry, density and tamped density, viscosity, turbidity, refractive index, light-scattering properties, and or sedimentation rate by ultracentrifugation. Silica particles are characterized by specific surface area as determined by the adsorption of nitrogen gas by the BET method (17) or by the adsorption of cetyltrimethylammonium bromide (CTAB) (18); average particle size and size distribution; sieve residue; porosity such as average pore diameter and pore volume as determined by intrusion of mercury or adsorption of nitrogen gas (19); degree of aggregation as determined by mean projected area (MPA) measured using transmission electron microscopy (20); oil absorption such as dibutylphthalate, dioctylphthalate, or linseed oil; and rate of dissolution. Particle size is measured by transmission electron microscopy (21), light scattering, absorbance of visible light, low angle x-ray scattering, centrifugation, rate of reaction with molybdic acid, or surface area measurement of dried particles. A summary of standard test methods used to characterize amorphous silica follows:

Property tested	Test methods
color	ISO 787, Parts I, XII; ISO 5794, Part I
average particle size	ASTM C721, D1366, E20, E523, and F660; ISO 787, Part XVIII
SiO ₂ content	ASTM C575 and D297; ISO 5794, Part I
coarse, insoluble material	ASTM C117 and D185; ISO 787, Parts III, VI, VII, and VIII; ISO 5794, Part I
volatile matter (water)	ISO 787, Part II
specific surface area	ASTM D1993 and D5604; ISO 5794, Part I; NF T 45-007
bulk density	ASTM D604; ISO 787, Parts X, XI
mean projected area of aggregates	ASTM D3894
metal ion and salt content	ASTM C575 and D719; ISO 5794, Part I, Annexes A, B, C, D
absorption of oil	ASTM D2414; ISO 787, Part V
pH	ASTM D1512; ISO 787, Parts IV, IX
porosimetry	ASTM D4284
occupational exposure	ASTM E1156

The typical range of properties for commercial colloidal silicas, silica gels, precipitated silicas, and pyrogenic silicas is given in Table 1.

Table 1. Properties of Commercial Amorphous Silica

Property	Colloidaln silica	Silica gels ^a	Precipitated silica	Pyrogenic silica
SiO ₂ , %	15–50	96.5–99.6	85–95	98.3–99.8
surface area, m ² /g	50–750	200–800	25–700	35–410
oil absorption, g /g	0.9–3	1.5	1.5–3.5	0.5–3
pH, aqueous	3–5, 8–11	2.3–7.4	5–9	3.5–5
weight loss, % at 105°C	50–80		4–7	0.5–2.5
at 1200°C	50–90	2–17.5	10–14	1–4
density, g /cm ³	2.2–2.3	2.22	1.9–2.1	2.16
bulk density, g /cm ³	1.2–1.4	0.1–0.9	0.03–3	0.03–0.12
particle size ultimate, nm	4–60	1–100	5–50	5–50
aggregate, nm			100–500	100–1000
agglomerate, μm		3–25	1–50	1–3
refractive index, n _D	1.35–1.45	1.35–1.45	1.45	1.45
Na ₂ O, maximum %	1	1	2	0.2

^a Dry.

Commercial Production

Approximately 40% of synthetic amorphous silica production is in Europe, followed by North America at 30%, and Japan at 12%. Although deposits of naturally occurring amorphous silicas are found in all areas of the world, the most significant commercial exploitation is of diatomaceous earth in industrialized countries (see DIATOMITE). This is because of the high cost of transportation relative to the cost of the material. Woddwide manufacturers of amorphous silica products are listed in Table 2.

Table 2. Principal Manufacturers of Amorphous Silicas

Producer ^a	Silica product	Manufacturing location	Capacity ^b , t × 10 ³
Akzo	precipitated, gel, colloidal	Germany, Netherlands, Thailand	85
Bayer	colloidal	Germany	6
Cabot Corp.	fumed	United States, Germany, United Kingdom	32
C-E Minerals	fused	United States	25
Chuo Denko, Denki Kagaku Kogyo, NKK, Showa Denko, and Toshiba Ceramics	fused	Japan	48 ^c
Crosfield	precipitated, gel	United Kingdom, United States	32
DeGussa	precipitated, fumed, colloidal	Belgium, Germany, India, Japan, Spain, United States	277
DuPont Chemicals	colloidal	United States	7
Eka Nobel	colloidal	Sweden, United States	8
General Electric	fumed	United States	7
J. M. Huber	precipitated	United States, Finland, India	130
Minco	fused	United States	21
Nalco Chemical	colloidal	United States, Europe	10
Nippon Silica	precipitated	Japan	18

PPG Industries	precipitated	United States, China, Netherlands, Taiwan, Thailand	210
PQ Corp.	gel, colloidal, precipitated	North America, Latin America	20
Precision Electro Minerals	fused	United States	9
Rhône-Poulenc SA	precipitated, gel	France, Korea, India, United States	87
Shionogi	precipitated	Japan	10
Solvay	gel	Germany	5
Tizayuca	precipitated	Mexico	8
Tokuyama Soda	precipitated, fumed	Japan	36
Wacker-Chemie	fumed	Germany, United States	25
W. R. Grace	gel	United States, Germany	29
others	gel, colloidal, precipitated		60

^a Includes joint venture companies.
^b May not include captive capacity for intermediates.
^c Total production capacity.

Heterogeneous Reactions

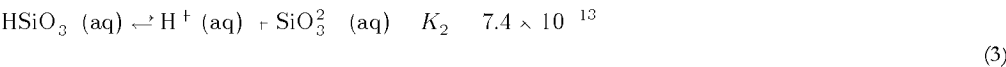
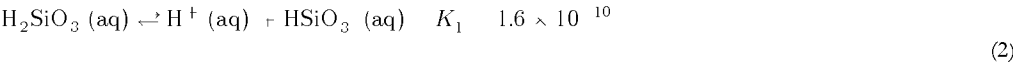
Dissolution. Amorphous silica dissolves or depolymerizes in water according to equation 1:



H₄SiO₄ (aq), which is either a monosilicic or an orthosilicic acid, is also expressed as Si(OH)₄ (aq) or H₂SiO₃ (aq). This material has not been isolated (1).

The solubility of amorphous silica in water at 25°C ranges from 70 to 150 ppm SiO₂ (1.2–2.2 mmol kg). The variation results from differences in particle size, degree of hydration, and level of trace impurities. Areas having a small positive radius of curvature (convex) dissolve most extensively, and areas having a small negative radius of curvature (concave) dissolve least extensively (22). Particles having a diameter of less than approximately 5 nm have a progressively greater solubility than do particles of continually increasing diameters. Particles <3 nm in diameter are very soluble (1). The solubility of amorphous silica in water increases with increasing temperature and pressure (23). For example, the solubility at 101.3 kPa (1 atm) and 0°C is approximately 50 ppm of SiO₂ (0.8 mmol kg), increasing to 100 ppm (1.7 mmol kg) at 25°C, and 750 ppm (12.5 mmol kg) at 100°C. The solubility increases by approximately 50% as pressure is increased from 101.3 kPa to 101 MPa (1–1000 atm) at temperatures between 0–25°C.

Hydrated amorphous silica dissolves more rapidly than does the anhydrous amorphous silica. The solubility in neutral dilute aqueous salt solutions is only slightly less than in pure water. The presence of dissolved salts increases the rate of dissolution in neutral solution. Trace amounts of impurities, especially aluminum or iron (24,25), cause a decrease in solubility. Acid cleaning of impure silica to remove metal ions increases its solubility. The dissolution of amorphous silica is significantly accelerated by hydroxyl ion at high pH values and by hydrofluoric acid at low pH values (1). Dissolution follows first-order kinetic behavior and is dependent on the equilibria shown in equations 2 and 3. Below a pH value of 9, the solubility of amorphous silica is independent of pH. Above pH 9, the solubility of amorphous silica increases because of increased ionization of monosilicic acid.



Amorphous silica is essentially insoluble in methanol, thus solubility in methanol–water mixtures decreases with increasing percentages of methanol. Natural biogenic silica often has an organic coating that inhibits dissolution. The coating can be removed by acid washing. The point of zero charge for the silica surface is between pH 2.5–3 (26).

Polymerization. Dissolved silica undergoes polymerization to give discrete particles that can further react. The three stages of reaction are polymerization of monomers to form particles, growth of particles, and linking of particles to give chains and then networks. In basic solution, the particles grow in size and decrease in number. In acid solution or in the presence of flocculating salts, particles aggregate into three-dimensional networks and form gels (Fig. 10) (27).

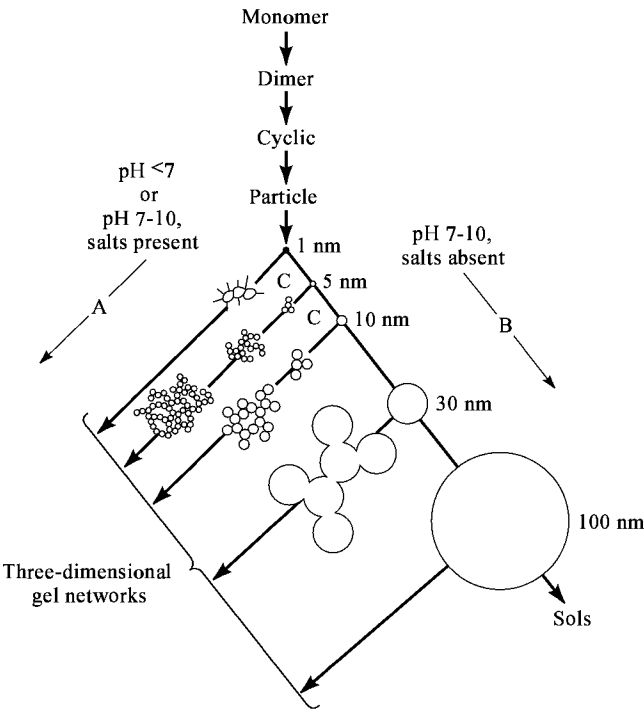


Fig. 10. Polymerization behavior of silica. In basic solution (B), particles grow in size and decrease in number; in acidic solution or in the presence of flocculating salts (A), particles aggregate into three-dimensional networks and form gels (1).

At concentrations below approximately 100 ppm of SiO₂, the solubility of amorphous silica in water is such that monosilicic acid is in a true solution in water. Polymerization occurs only when solubility is exceeded and there is no solid phase present on which the silica can be deposited. The polymerization is an ionic reaction proportional to the concentration of hydroxyl ion at pH > 2. Because of the thermodynamic tendency to maximize siloxane, -Si-O-Si-, and to minimize silanol (-Si-O-H) bonds, initially formed low molecular weight species are condensed, forming ring structures such as cyclic tetramer, hexamer (see Fig. 2a), and octamer (see Fig. 2b). These species continue to react with additional monomer and become linked to one another (see Fig. 3). Internal condensation reactions continue to occur that lead to formation of a small particle. At temperatures >80°C and pH values >7, the particle is anhydrous. The silanol groups generally only form on the surface. These particles serve as nuclei on which three-dimensional particle growth occurs. Because of the solubility difference, small particles dissolve and the silica is deposited on larger particles that grow in average size. This process is known as Ostwald ripening. Increasing temperature either increases the growth of sol particles or causes gel formation. The presence of dissolved salts assists in neutralizing particle surface charges to enhance aggregation of the particles into three-dimensional gels. The rate of aggregation also increases with dissolved silica concentration in solution. As polymerization occurs, the solution still contains dissolved silica at a concentration equivalent to the solubility of amorphous silica (1,27-29).

Silica Sols and Colloidal Silica

Properties. Colloidal silica is a stable aqueous dispersion or sol of discrete amorphous silica particles having diameters of 1 to 100 nm. Silica sols do not gel or settle out of solution for at least several years of storage. Aqueous sols that contain up to 50% silica have been developed (30,31). Particle sizes of approximately 130 nm in diameter are possible (32), but slowly settle out of solution.

In the absence of a suitable solid phase for deposition and in supersaturated solutions of pH values from 7 to 10, monosilicic acid polymerizes to form discrete particles. Electrostatic repulsion of the particles prevents aggregation if the concentration of electrolyte is below ca 0.2 N. The particle size that can be attained is dependent on the temperature. Particle size increases significantly with increasing temperature. For example, particles of 4-8 nm in diameter are obtained at 50-100°C, whereas particles of up to 150 nm in diameter are formed at 350°C in an autoclave. However, the size of the particles obtained in an autoclave is limited by the conversion of amorphous silica to quartz at high temperatures. Particle size influences the stability of the sol because particles <7 nm in diameter tend to grow spontaneously in storage, which may affect the sol properties. However, sols can be stabilized by the addition of sufficient alkali (1,33).

The stability of a silica sol depends on several factors (Fig. 11) and can be controlled by the surface properties of the silica (1). The pH value should be above 7 to maintain sufficient negative charge on the silica particle surface to prevent aggregation. This surface charge is neutralized by soluble salts that ionize and reduce the size of the double layer around the silica surface (Fig. 12), allowing aggregation. Therefore, sols are only stable at low salt concentrations. In the low pH value region sols are metastable, and gelling and aggregation are catalyzed by even very small amounts of fluoride ion. In this low pH value region, water-miscible organic solvents, like alcohol, retard gelling. Gelling occurs more rapidly at higher temperatures and is facilitated by a high silica concentration in the sol.

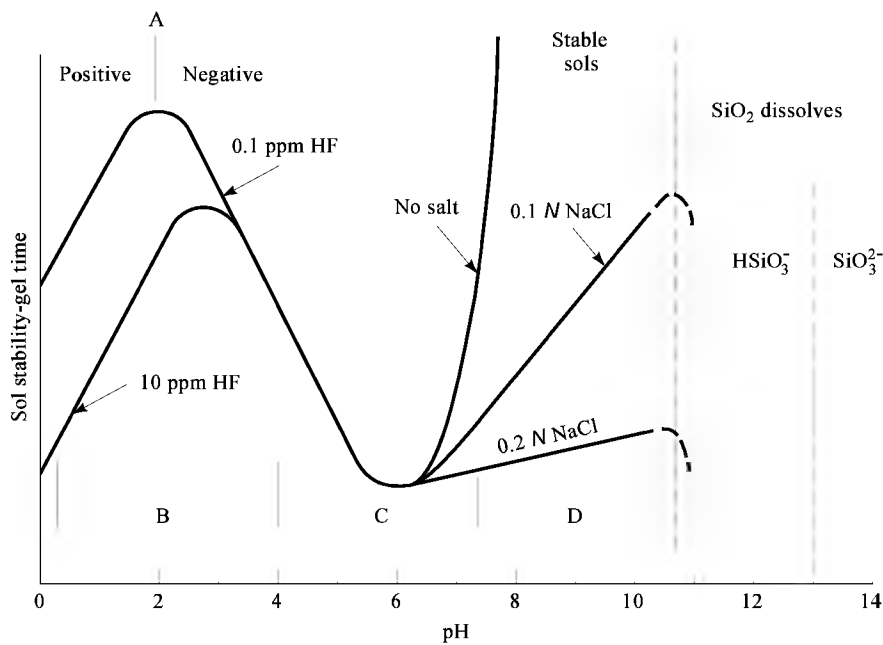


Fig. 11. Effects of pH in the colloidal silica-water system (1), where A represents the point of zero charge; regions B, C, and D correspond to metastable gels, rapid aggregation, and particle growth, respectively. Positive and negative correspond to the charges on the surface of the silica particle.

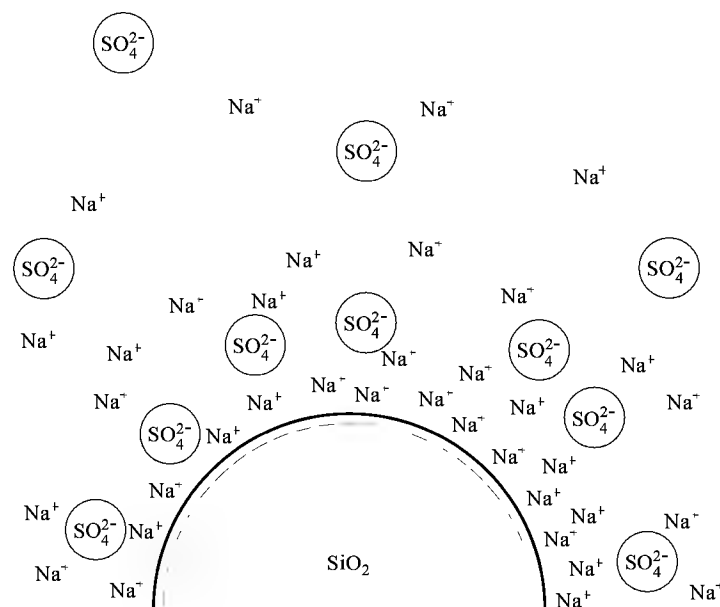


Fig. 12. Salt retention by colloidal particles. The curved dashed and solid lines represent the surface of a negatively charged silica particle. Around this there is a layer of counter sodium cations; outside there is a layer in which sulfate anions (○) are more concentrated than in the bulk solution.

Silica sols can be destabilized by aggregation, gelation, crystallization, or particle growth plus settling. Aggregation occurs by coagulation in which particles collide or by flocculation in which particles become linked by bridges of flocculating agent (see FLOCCULATION AGENTS). Aggregation is prevented at neutral or lower pH values by the presence of an adsorbed layer of inert material, such as a surfactant, on the silica surface, thus preventing the direct contact of silanol groups with one another. Aggregation, as well as particle growth, is minimized by maintaining either a low silica concentration, a low reaction temperature, and a pH value of 9–10. Destabilization of a sol by crystallization of silica rarely occurs because this is a slow process at ambient temperatures (1).

Preparation. To produce sols that are stable at relatively high silica concentrations, particles must be grown to a certain size in weakly alkaline aqueous dispersion. Sols can be prepared by (1) neutralizing a dilute solution of sodium silicate with acid to a pH value of 8–9. A sol is produced if the sodium salt concentration is <ca 0.3 N and the neutralization reaction is performed at elevated temperatures to promote rapid particle growth. Rapid mixing prevents formation of low pH regions where gelation might occur; (2) washing sodium ions from the sodium silicate solution by electrodialysis or using an ion-exchange resin to form active silica, ie, low molecular weight polysilicic acid (1), which is stored at 82–100°C. Additional active silica is added to grow the sol particles, which can then be concentrated by evaporation or ultrafiltration; (3) washing out the salt from a freshly formed silica gel and peptizing at elevated temperature and pressure; (4) hydrolyzing SiCl_4 or organosilicates, such as ethyl silicate plus isopropyl alcohol. However, such sols are generally not stable and gel quickly; (5) cleaning pulverized silicon metal with hydrofluoric acid to remove the oxide coating, which causes a reaction with water in an alkaline medium to produce hydrogen gas and colloidal silica, which can be stabilized as a sol with alkali; and (6) oxidizing SiCl_4 or ethyl silicate at high temperature, or vaporizing silica in the presence of a reducing agent to give pyrogenic silica powder. Particles are partly coalesced, have a low concentration of surface silanol groups, and are difficult to disperse in water to form sols without the addition of a wetting or other dispersing agent, or such treatment as grinding. The resultant sols contain aggregates. A detailed review of silica sol preparation methods is available (1).

Silica sols can be purified with the aid of an ion-exchange resin, dialysis, electrodialysis, or washing. Washing is usually affected by centrifugation, flocculation, ultrafiltration, or electrodecanting, followed by redilution and reconcentration. A preservative such as formaldehyde is sometimes added to prevent the growth of microorganisms. Sols are concentrated by evaporation using forced-circulation evaporators. If the sol becomes too concentrated, a hard layer of silica is deposited on the equipment walls, especially on heat-exchanger surfaces, so evaporation must be carefully monitored. Sols having particle size >30 nm in diameter are concentrated by centrifugation. Ultrafiltration (qv), where the sol collects near the impermeable filter, can be used to remove water and salts. Sols that are low in soluble salts are concentrated by electrodecantation (1).

Modifications. Coagulation, flocculation, or gelation of silica sols and their subsequent drying produce amorphous silica powders. Gelation is rapid at pH values of 4 to 6. It is effected by increasing the silica content, such as by evaporation (qv). Further drying of the resulting gel yields colloidal aggregates. Coagulation of a sol is prevented by surface charge and surface hydration; conversely, neutralizing the surface charge by lowering the pH value or by adding salt causes coagulation. Addition of salts also causes surface adsorption of cations that reduce surface hydration. Polyvalent cations such as Al^{3+} are strongly adsorbed by silica surfaces and neutralize the surface charge. Hydrated metal cations act as bridging compounds or flocculating agents (qv) for silica particles. Flocculating agents used with silica sols are cationic surfactants (qv) such as octadecyltrimethylammonium bromide or organic polymers such as polyethyleneimine or polyacrylamide (1).

Sol particles are extremely small, and particulate silica obtained from sols is composed of aggregates or porous particles that have a much higher specific surface area than estimated from apparent size. Aggregates are also called secondary particles. Once silica particles have been formed as a sol, the surface can be modified by the attachment of different atoms or groups to obtain specific properties. Addition of aluminate ions gives an aluminosilicate surface that is much less soluble than is the silica surface and much more stable toward gelling, even at pH values of 4 to 6 and in the presence of salts. Addition of polyvalent metal oxides reverses the surface charge of silica and prevents bridging by siloxane bond formation. Organic groups can be added to silica surfaces in the form of organic ions through the formation of $\text{Si}-\text{O}-\text{C}$ bonds as in esters or $\text{Si}-\text{C}$ bonds as in organosilicon compounds. These surface alterations make silica easily dispersible in organic solvents, including nonpolar hydrocarbons. Some of these processes may cause the silica surface to become hydrophobic. Even a small change in particle surface characteristics may change the chemical behavior of the particle. Thus, a silica particle covered with alumina behaves in many ways as an alumina particle, and a silica particle covered with a hydrocarbon coating acts in many ways as a large hydrocarbon molecule (1,34–37).

Applications. In 1994, approximately 80,000 metric tons of colloidal silica (38) were manufactured by DuPont, Eka Nobel, Nalco, and PQ (Custom Colloids, Nyacol Products) in the United States; by Bayer (Germany), Eka Nobel (Sweden), Monsanto (United Kingdom) in Europe; and by Nissan in Japan. Colloidal silicas are typically used for making silica gels having uniform pore sizes and pore volumes; as binders in molds for precision casting of superalloys; for increasing the friction properties of surfaces such as paper, floors, fibers, and films; as antisoiling finishes on paper, textiles, and painted surfaces; for hydrophobizing surfaces and increasing their wettability by water; for strengthening polymer and latex goods; for polishing silicon wafers used in electronics industry; for improving performance of agents used for wetting and dispersing; as photographic emulsions; and for clarifying wines (see WINE), beers (see BEER), and gelatin (qv) (39).

Silica Gel

Properties. Silica gel (see Fig. 8) is a coherent, rigid, continuous three-dimensional network of spherical particles of colloidal silica. Both siloxane, $-\text{Si}-\text{O}-\text{Si}-$, and silanol, $-\text{Si}-\text{O}-\text{H}$, bonds are present in the gel structure. The pores are interconnected and filled with water and/or alcohol from the hydrolysis and condensation reactions (40). A hydrogel is a gel in which the pores are filled with water. A xerogel is a gel from which the liquid medium

has been removed, causing the structure to collapse, thus decreasing porosity. If the liquid medium is removed in a manner such that shrinkage and collapse of the gel structure do not occur, an aerogel is formed. Porous glass is similar to a silica gel. Silica powder can be made from xerogels by grinding or micronizing, which decreases the size of the gel fragments but leaves the ultimate gel structure unchanged. Gels and powders are characterized by the density, size, and shape of the particles, particle distribution, and aggregate strength or coalescence (1,27–29,41–44).

Silica gels are classified into three types: regular, intermediate, and low density gels. Regular density silica gel is made in an acid medium, which gives high (eg, 750 m² g) surface area, small ultimate (or primary) particles having average pore diameters of 2.2–2.6 nm, and a pore volume of 0.37–0.40 mL g. This regular density silica gel contains approximately 6 wt % water as surface hydroxyl groups, which imparts a high capacity for adsorption of water and other polar molecules (see DESICCANTS). The gel exhibits a high selectivity for polar molecules and has a large percentage of small pores. Intermediate density silica gel consists of larger ultimate particles having a lower (300–350 m² g) surface area, larger (0.9–1.1 mL g) pore volumes, and larger (12–16 nm) average pore diameters. Because of the large pore size, intermediate density silica gel has a high capacity for water adsorption at high humidities. It is often used as a fine powder because aggregate (or secondary) particle size and porosity can be controlled. Low density silica gel (such as an aerogel) has lower (100–200 m² g) surface areas, larger (18–22 nm) average pore diameters, and larger (1.4–2.0 mL g) pore volumes. It is usually prepared as a very fine powder of extremely low density. Shrinkage of the gel during drying is minimized.

Preparation. Silica gels can be prepared by several methods (1,40,43). Most commonly, a sodium silicate solution is acidified to a pH value of <10–11. The gelling time varies, as shown in Figure 10. Syntheses include the bulk-set, slurry, and hydrolysis processes. The bulk-set process consists of preparing a silica hydrosol (gel) by mixing sodium silicate and a strong mineral acid, followed by mechanically breaking, washing the hydrogel particles free of electrolytes, drying, and activating. In the slurry process, sodium silicate solution and acid are mixed, either in batch or semicontinuously, to produce a gelatinous precipitate which is washed and dried, often by spray drying. These gels usually have a small particle size and the salts must be washed away before use. Gels can also be made directly from salt-free colloidal silica, which provides larger ultimate particle sizes and hence greater stability, along with low salt contents. Such gels may also have lower specific surface areas and larger pore diameters. Hydrolysis of pure silicon compounds such as ethyl silicate, silicon tetrachloride, and other volatile hydrolyzable silicic esters is another method of preparing gels. Although more expensive, this process produces dense gels of high purity and very small pore size.

The properties of a finished gel are determined by the size of the primary particles at the moment they aggregate into the gel network, called the gelation point (or gelation time). It is usually defined as the point at which aggregation in the sol forms a three-dimensional network that can support a stress elastically. The sol-to-gel transformation is not well defined because it occurs as a continuous increase in viscosity (40). The concentration of the primary particles in solution affects the physical properties of the gel, including the compactness of the gel network, the pH, salt concentration, temperature, and time during which the gel is aged while wet, and the mechanical pressure or shear forces applied to the gel before or during drying. Additionally, the physical properties are influenced by the temperature, pressure, pH, salt content, and surface tension of the liquid medium as it is being evaporated from the pores of the gel, and the temperature, time, and type of atmosphere in which the gel is heated after being dried (1). The type of catalyst used in gelation can have significant effects on the kinetics of the process and the structure of the silica gels formed (45,46). For some applications, silica gel is converted to pelletized or granular form by extruding pulverized gel with a binder or by shaping the hydrogel during drying. Silica can be gelled in spherical form by spray-drying, or by spraying droplets into an immiscible liquid (emulsion polymerization). Freezing of a silica sol produces silica–gel particles of nonspherical shapes.

Characterization. When silica gel is used as an adsorbent, the pore structure determines the gel adsorption capacity. Pores are characterized by specific surface area, specific pore volume (total volume of pores per gram of solid), average pore diameter, pore size distribution, and the degree to which entrance to larger pores is restricted by smaller pores. These parameters are derived from measuring vapor adsorption isotherms, mercury intrusion, low angle x-ray scattering, electron microscopy, gas permeability, ion or molecule exclusion, or the volume of imbibed liquid (1).

Surfaces can be categorized as fully hydroxylated, in which the surface consists solely of silanol, –Si–O–H, groups, a siloxane, –Si–O–Si–, or an organic surface. Silanol surfaces are formed by drying silica gels or precipitates from water at temperatures below 150°C. These surfaces are readily wetted by water. Hydroxylated surfaces heated from 300–1000°C progressively develop a siloxane surface by dehydration such as that found on pyrogenic silica surfaces. The behavior of particles having organic surfaces depends on the coating material. The particles may become dispersible in water, oil, or other organic solvents. If fluorocarbons are the surface group, the silica becomes both hydrophobic and oleophobic. The nature of the surface can be determined by measuring the heat of nitrogen adsorption, dye adsorption, infrared adsorption, or chemical analysis (1,28,33–37).

Modifications. Once a gel structure is formed, it can be modified in the wet state to strengthen the structure or enlarge the pore size and reduce surface area, through a process called aging (1,40,47). Silica gel reinforcement can be carried out in several ways: (1) active silica, ie, low molecular weight polysilicic acids (1), can be added to a broken-up gel in order to deposit it at a uniform rate on the gel; (2) active silica can be added to a sol as the gel is growing, causing strong gel bridges to form between particles; (3) wet gel can be heat-aged to increase coalescence of the particles.

In this aging process, silica is dissolved from smaller particles and deposited at the points of contact between larger particles. Condensation reactions in the gels continue to occur because of the presence of the silanol, –Si–O–H, groups (48,49). Because of the formation of new siloxane bonds that bridge the particles, the continued polymerization reactions strengthen and stiffen the network (47). Washing can also be an aging step because aging a wet gel can increase interparticle bonding, which leads to less shrinkage of the gel during drying. Low density silica gels are made by minimizing shrinkage during drying by reinforcing the gel by aging, or by replacing the water with a liquid of lower surface tension, for example an alcohol. The gel is then heated to a temperature above the critical point of the liquid, thereby releasing the liquid as a vapor. Supercritical drying is often used in aerogel processing (1).

Sintering, or heating to convert a powder into a continuous mass, a dried silica surface in air or in a vacuum causes shrinkage that decreases the surface area, whereas sintering in steam increases pore size. For example, micropores are obtained by heating a hydrated gel at 1000°C for 10 hours. The presence of impurities such as aluminum tends to minimize changes caused by heating. At temperatures >1000°C, however, and in the presence of impurities, silica gel is converted to cristobalite or nonporous silica glass. Gels having extremely small pores can be made. Such gels include impervious silica, porous glass, and adsorbents for which the applications are tailored to surface properties and pore size.

Applications. Approximately 80,000 metric tons of silica gels were manufactured in 1994 (38), primarily by Crosfield, Grace, PQ, and SCM in the United States; by CECA (France), Chemische Fabriken Oker und Braunschweig (Germany), Grace (Germany), and Solvay (Germany) in Europe; and by Dohkai, Mizusawa, and Tokai in Japan. The largest application areas for silica gels include health care products as thixotropic agents in cosmetics (qv) and dentrifices (qv); selective absorbents to maintain clarity during the brewing of beer; desiccants; thixotrope and flattening agents in coatings (50); and supports for polymerization catalysts. Specialized silica gels that have uniform pore structure, and which are often surface-treated, are used in the production of chromatography (qv) columns for gas- and liquid-phase separations. Amorphous silica gels (51) and aerogels (52) are also used as insulating materials and sol–gel coatings for optical applications (53).

Precipitated Silica

Particulate silica is composed of aggregates (or secondary particles) of ultimate (or primary) particles of collidal-size silica that have not become linked in a massive gel network during the preparation process. Particulate silicas are either made from the vapor phase to form pyrogenic (or fumed) silicas, or by precipitation from solution, generally aqueous, to form precipitated silicas. Particulate silica powders have a more open structure with higher pore volume than do dried pulverized gels, as can be seen by comparing Figures 6 and 9 to Figure 8.

Properties. The physical and chemical properties of precipitated silicas can vary according to the manufacturing process. Ultimate particle and aggregate sizes of silicas precipitated from solution can be varied by reinforcement and control of suspension pH, temperature, and salt content value. The surface area, as determined by nitrogen (17) or CTAB (18) adsorption, is a function of the ultimate particle size, which can range from 5 to 50 nm in diameter. Aggregates are three-dimensional clusters of ultimate particles (see Fig. 9), which range in size up to approximately 500 nm in diameter. Ultimate particles in aggregates are covalently bonded to one another via siloxane bond formation. Aggregate particles can be physically agglomerated through hydrogen bonding of surface silanol groups to one another, thus affording structures that can range up to approximately 100 μm in diameter. The median agglomerate particle size is generally 20 to 50 μm in diameter, but can be reduced in size by milling, for example, to approximately 1 μm.

Preparation. Precipitated silica is formed from an alkaline metal silicate solution, such as sodium silicate in a ratio of approximately 2.5–3.3

$\text{SiO}_2\cdot\text{Na}$. Lower concentrations of silicate are used in silica gel preparation. In the absence of a coagulant, silica is not precipitated from solution at any pH value (1). Silica is precipitated by adding acid to sodium silicate to reduce the pH value of the hot suspension to a pH value of 9–10, where the concentration of sodium ion exceeds approximately 0.3 *N*. Sulfuric acid, hydrochloric acid, and carbon dioxide are examples of acids used commercially to neutralize sodium silicate and precipitate silica. Figure 13 shows a schematic of this process. Precipitation proceeds in three basic steps: forming colloidal particles through nucleation and particle growth to the desired ultimate particle size; coagulating colloids into aggregates to form a suspended precipitate, where the control of pH and sodium ion concentration is critical; and reinforcement of the aggregate particle to the desired degree without further nucleation (1) so that the structure is not altered.

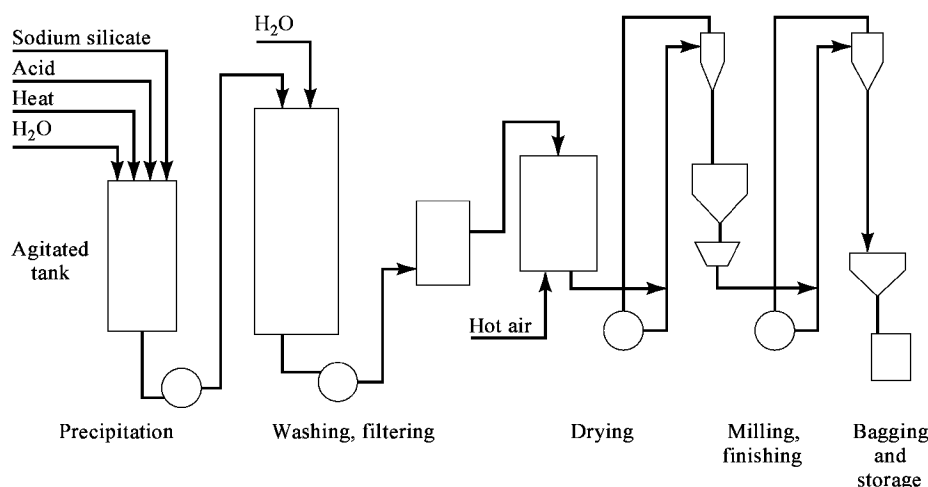


Fig. 13. Schematic of precipitated silica production (54), where $\square$, eg, reactor, washer, filter, and dryer; $\bigcirc$ pump or transfer device, eg, bucket elevator and air conveyor; ∇ storage bin; ∇ optional size modifier, eg, mill and granulator.

Coagulation can also be induced by using polyvalent metal cations such as calcium ions, as well as by using ammonium and fluoride ions, and certain organic compounds. Reinforcement can be carried out by adding active silica, low molecular weight polysilicic acids, to the aggregate suspension under alkaline conditions above 60°C, or by adding acid and metal silicate at a controlled rate so that active silica forms and is polymerized on the suspended silica without forming new particles. To obtain a dry precipitate by water removal, the ultimate particles should be larger than approximately 10 nm, typically >20 nm, and the aggregates should be reinforced so that the pore structure does not collapse upon drying as a result of capillary forces. This reinforcement of the silica aggregate is characterized by the buildup ratio, which is the ratio of the final weight of silica in the system to the weight of the aggregated silica. Buildup ratios of less than 4:1 are usually preferred in order to achieve, for example, dispersibility in rubber. For each type of precipitated silica there is an optimum degree of reinforcement for maximum dispersibility (1).

Modifications. Precipitated silicas are hydrophilic and may contain up to approximately 10% water as surface silanol, $-\text{Si}-\text{O}-\text{H}$, groups that remain after drying at 150°C. The interaction of the surface of amorphous silicas with organic molecules is strongly affected by the degree of hydroxylation of the silica surface. Thus, modification to hydrophobize the surface of a precipitated silica can be an important step toward increasing its compatibility with substrates such as nonpolar polymers. The choice of appropriate modifying substance and the method of surface modification are important. Surfactant (55–57), organosilane (58), and organotitanate (59) coupling agents and polymeric films (60,61) may be employed.

Applications. In 1994, approximately 675,000 metric tons of amorphous precipitated silica were manufactured for sale (38,62). Degussa, J. M. Huber, and PPG in the United States and in Europe, and Akzo (Germany), Alufloor (Sweden), Crosfield (United Kingdom), Nippon Silica (Japan), Rhodne-Poulenc (France), Shionogi (Japan), Tokuyama Soda (Japan), and Vitro PQ (Mexico) are the primary producers.

Additionally, precipitated silica is manufactured and utilized internally for such uses as extenders for detergents and cleaners. Because of the small particle size and complex aggregate structure (see Fig. 9), precipitated silica imparts the highest degree of reinforcement to elastomer compounds, eg, natural rubber, among all of the mineral fillers (see RUBBER, NATURAL). This superior reinforcement is employed in a variety of rubber compounds for shoe soles, industrial rubber goods, and tires (63). Precipitated silica is used in shoe soles for its resistance to wear and tearing, for its nonscuffing characteristics, and in order to obtain compounds having light color, or even a transparent material. In addition, precipitated silica is used to improve the tear strength and resistance to flex fatigue, ie, cracking and cut growth, and heat-aging in a wide variety of manufactured rubber goods, including conveyor and power transmission belts, hoses, motor and dock mounts, and bumper pads. Rubber rolls that utilize the abrasion resistance, stiffness, and nonmarking characteristics of precipitated silica are important for use in paper processing and the dehulling of grains, particularly rice (see WHEAT AND OTHER CEREAL GRAINS) (64).

Precipitated silica is used in a variety of tire applications (Fig. 14). The reinforcement of rubber compounds by precipitated silica has been found to be directly related to the surface area of the silica (65,66), which is a function of the ultimate particle size. The use of precipitated silica at very high levels in tire treads has been shown to impart outstanding wet traction and rolling resistance to passenger tires, which improve vehicle handling and fuel economy (67–69). Newer grades of precipitated silica having smaller ultimate particle sizes and intricate aggregate structure (70) have shown exceptionally high reinforcement of tire tread compounds (71) and the potential for improved treadwear in tires.

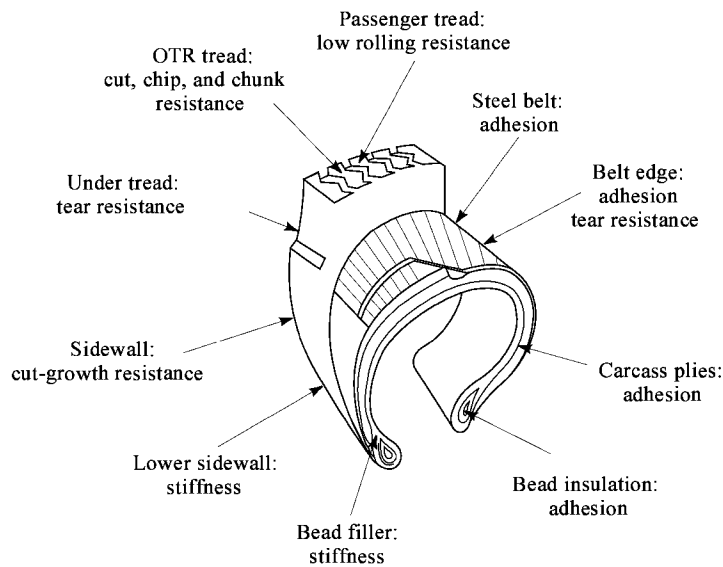


Fig. 14. Schematic of a passenger tire showing individual components and use of precipitated silica to improve rubber compound performance. OTR off-the-road.

Additional uses of amorphous precipitated silica include reinforcement of specialty elastomers, especially those based on polydimethylsiloxane (72); defoamer in paper processing (73) and an ingredient for good ink retention; production of polyethylene silica membranes used as separators for plates in lead–acid batteries (qv) (74,75); control of viscosity in dentifrice (76) and coating formulations (77); abrasive ingredient in toothpaste (76) and polishing formulations; absorbent of liquids in food applications (78), animal feedstocks, and fire extinguishers; starting material for high purity zeolite manufacture; gloss control agent in coatings and plastics; carrier for active ingredients such as pesticides (79); and extender for detergents, papers, and coatings (54).

Pyrogenic Silica

Properties. Amorphous pyrogenic (fumed) silicas are fluffy white powders that are generally less dense and of higher purity than are silicas precipitated from solution. Pyrogenic silicas have a much lower hydrated surface and are sometimes completely anhydrous. Surface silanol density normally ranges from 2–4 nm² of surface area. The particle size is controlled by combustion conditions during the flame hydrolysis. Pyrogenic silicas having surface areas <300 m² g are essentially nonporous, whereas those having higher surface areas can have some porosity. They generally can contain a few hundred ultimate particles fused into branched-chain, three-dimensional aggregates (54) (eg, see Fig. 6). Pyrogenic silicas in the 100 nm to 2 μm diameter particle size range are common (80).

Preparation. Pyrogenic silica can be prepared in several ways, including vaporizing silica and oxidizing organic or inorganic silicon compounds. Vaporizing silica at high temperatures of approximately 2000°C forms anhydrous amorphous silica particles upon cooling. In the presence of a reducing agent such as coke, silica sublimates at about 1500°C to produce the volatile silicon monoxide, SiO, which can then be oxidized to produce particulate fumed silica, SiO₂. Oxidation of silicon tetrachloride, SiCl₄, vapor at high temperatures produces fumed silica and Cl₂. Alternatively, silicon tetrachloride can be burned in the presence of methane, CH₄, or hydrogen, H₂, gases to produce fumed silica, H₂O, and HCl. The latter process is an important commercial method (Fig. 15) (54). Silicon ester vapors can be oxidized and hydrolyzed to produce particulate silica of high purity, though at high cost. Silicon tetrafluoride, SiF₄, a by-product of the phosphate fertilizer industry, can also be used to produce pyrogenic silica by hydrolysis of the vapor at 1600–2200°C. HF is a by-product that can then react with sand to produce more SiF₄ (1,81).

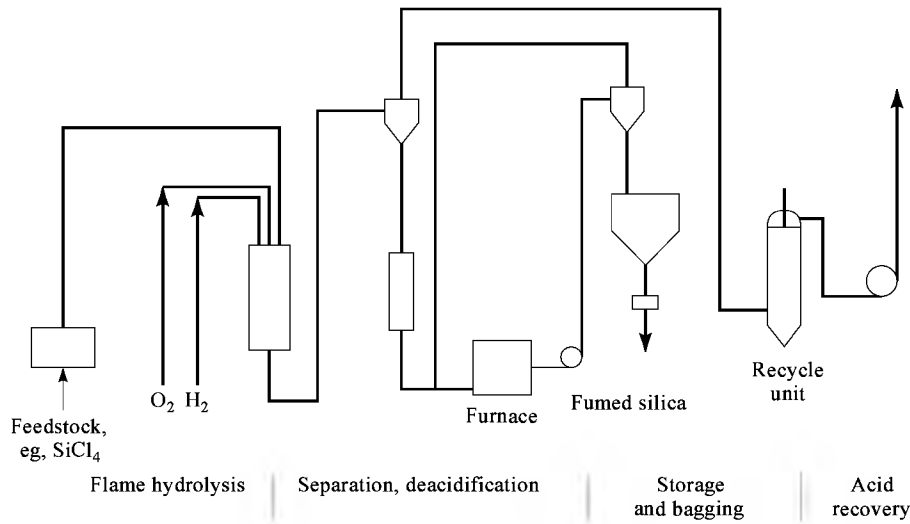


Fig. 15. Schematic of pyrogenic silica production (54). See Figure 13 for definitions.

Applications. Pyrogenic or fumed silica production in 1994 was estimated to be 140,000 metric tons (38). Cabot and Degussa in the United States and in Europe, General Electric (United States), Nippon Aerosil (Japan), and Tokuyama Soda (Japan) were the largest manufacturers. Pyrogenic silica is used in rubber applications that require a low level of surface water per unit surface area of the silica, primarily as a reinforcement in adhesives, sealants (82), and elastomer compounds based on silicone polymers (83,84). Pyrogenic silica also has a variety of other uses. It serves as a thixotropic agent in plastics, eg, in polyester casting resins; gel coats (85); sealants and adhesives (86); cosmetics in ointments and lotions (87); toothpastes; coatings (85); printing inks (88); free flow, antistatic agent in animal feedstuffs and hygroscopic powders; carrier for active ingredients such as pharmaceuticals (qv), catalysts, and liquid spice extracts (89); gloss control agent in coatings and foils (54); coating additive to improve the mechanical reliability of fused silica optical fibers (90); and antifoaming or defoaming agent in the manufacture of paper, decaffeinated coffee and tea, poultry and seafood processing, and oil refining (91).

Fused Silica

Transparent fused silica can be formed at a temperature of 1200°C and a pressure of 13.8 MPa (2000 psi) from silica powder consisting of 15 nm ultimate particles (92) or by electric arc fusion of pure silica sand having low iron and alkali metal contents. The cooled product is ground to the desired particle size. Fused silica is primarily manufactured by C-E Minerals, Minco, and Precision Electro Minerals in the United States; by Chuo Denko, Denki Kagaku Kogyo, NKK, Showa Denko, and Toshiba Ceramics in Japan. Based on 1988 data and projected growth, an estimated 135,000 metric tons of fused silica were used in 1994 as a sacrificial component or investment casting in the manufacture of metals and as a component in refractory materials (62).

Naturally Occurring Amorphous Silica

Formation. Most naturally occurring amorphous silica deposits are biogenic forms deposited from the exoskeleton, plates, or spines of aquatic organisms such as diatoms, radiolarians, silicoflagellates, and certain sponges. These organisms extract silica from very dilute solutions, eg, 0.1 ppm of SiO₂ or 2 mmol kg, and are found in both marine and freshwater environments. Large deposits of these materials are found as loosely coherent chalk-like sediments in the equatorial Pacific area, ie, radiolarians, and in high latitude areas of all oceans, ie, diatoms. All biogenic silicas are noncrystalline because the kinetic barrier to crystallization is high (ca 800 kJ mol (191 kcal mol)) in neutral aqueous environments. In all these structures the mineral exists as a hydrated, covalent inorganic polymer having the general structure (SiO_{n-2}(OH)_{4-n})_m, where *m* is a large integer representing the degree of polymerization and *n* ranges from 0–4 to indicate the degree of hydration (93).

The silica structure has a close relationship to the cellular structure of the organism. Silica is deposited as gels, sheets, fibers, tubes, and globular assemblies. It has been found that the macroscopic geometric structure of the deposited silica is replicated in the short-order (two- to three-siloxane unit) structure (93). Although the dissolution of silica is inhibited by inclusion of various metal ions in the structure and often an organic coating, these deposits are converted over millions of years to opal-CT and opal-C, which are well-crystallized cristobalite (94). These siliceous materials are also transformed by dissolution, precipitation, recrystallization, and compaction over geological time to form a less soluble, denser material, which is called chert if granular, chalcedony if fibrous, or flint if dark gray. Amorphous silica can be formed by the alteration of sand to a colloiddally subdivided high surface area, amorphous silica volcanic ash (1). Amorphous silica is sometimes precipitated from the hot supersaturated waters of hot springs (siliceous sinter) and geysers (geyserite) where it is often found along with calcareous sinter (95,96).

Applications. The most significant commercial material from the biogenic silicas is diatomaceous earth [7631-86-9], also called kieselguhr or diatomite (qv). These deposits are the sediment of fragments and shells of one-celled algae, which form very fine particles having high surface area and as high as 94% silica content. Initially, silica in diatomite is amorphous, but portions of these deposits have been converted to cryptocrystalline quartz (95). Approximately 1,500,000 metric tons of diatomaceous earth were used in 1994 as absorbents, fillers, insulating materials, and polishing agents, primarily in China, France, Germany, Italy, Korea, Mexico, and the United States. Tripoli, a white-to-gray, soft, porous material used as an abrasive or a filler, is mined from the Devonian deposits in southern Illinois. It is sold as amorphous silica, although it is actually a microcrystalline quartz (<10 μm dia) formed by leaching of carbonates from calcite- or dolomite-bearing chert (see ABRASIVES; FILLERS).

Health and Safety Factors

Amorphous silica is classified as nontoxic by ingestion. Soluble silica is present in most drinking water. It has been observed that silica passes through the body in soluble form being excreted with the water (39).

Unlike crystalline silica, amorphous silica has not been shown to cause silicosis (97), even in workers having long exposure to the product. This is presumed to be a result of the differing surface chemistry and solubility between crystalline and amorphous silica (98). The biological effects in animals of inhaled amorphous precipitated, fumed, and gelled silicas have been studied (99). After 12 to 18 months exposure to an environment containing an average daily respirable dust concentration of 6.9–9.9 mg m⁻³, rats, guinea pigs, and Cynomolgus monkeys were tested for pulmonary function and by microscopic examination of internal organ tissue sections. Respiratory function was impaired for all animals and the presence of macrophage and cell aggregate lesions in the lungs of the monkeys were reported. It has been reported that these lesions regress after exposure to the dust is stopped (100). Monkeys exposed to fumed silica show the presence of reticulin fibers in the lungs as well.

Amorphous silica is classified by the Occupational Safety and Health Administration as a nuisance dust. The principal reported health reaction is contact dermatitis resulting from the absorption of protective oils from the skin (1).

BIBLIOGRAPHY

"Silica and Silicates (Silica Gel)" in *ECT* 1st ed., Vol. 12, pp. 345–360, by S. S. Hubbard, The Davison Chemical Corp.; "Silica (Amorphous)" in *ECT* 2nd ed., Vol. 18, pp. 61–72, by R. K. Maher, Davison Chemical Division, W. R. Grace & Co.; in *ECT* 3rd ed., Vol. 20, pp. 766–781, by J. D. Willey, University of North Carolina at Wilmington.

1. R. K. Iler, *The Chemistry of Silica*, John Wiley & Sons, New York, 1979.
2. W. Nieuwenkamp, *Z. Krist.* **96**, 454 (1937).
3. F. Ordway, *Science*, **143**, 800 (1964).
4. H. P. Klug and L. E. Alexander in *X-Ray Diffraction Procedures for Polycrystalline and Amorphous Materials*, 2nd ed., John Wiley & Sons, New York, 1974.
5. E. Gorlich, *Ceramics Int.* **8**, 3 (1982).
6. W. J. Miles and R. D. Hamilton, *Anal. Chim. Acta*, **286**, 3 (1994).
7. R. Schwartz and O. Liede, *Chem. Ber.* **53B**, 1680 (1920).
8. H. Kautsky and R. Irnich, *Z. Anorg. Allg. Chem.* **295**, 193, 206 (1958).
9. H. Kautsky, W. Otto, and H. Pfeleger, *Z. Naturforsch.* **19**, 102 (1964).
10. H. Kautsky and R. Reise, *Z. Naturforsch.* **20b**, 93 (1965).
11. E. Wiberg and W. Simmler, *Z. Anorg. Allg. Chem.* **283**, 401 (1956).
12. **U.S. Pat. 2,801,902** (Aug 6, 1957), G. B. Alexander and R. K. Iler (to E. I. du Pont de Nemours & Co., Inc.).
13. T. Nemetshek and U. Hofmann, *Z. Naturforsch.* **8b**, 410 (1953).
14. A. Simon and P. Rath, *Z. Anorg. Allg. Chem.* **202**, 191 (1931).
15. E. P. Bertin, *Principles and Practices of X-Ray Spectrometric Analysis*, 2nd ed., Plenum Press, New York, 1975.
16. A. T. Bell and M. L. Hair, eds., *Vibrational Spectroscopies for Adsorbed Species*, American Chemical Society, Washington, D.C., 1980.
17. S. Brunauer, P. H. Emmett, and E. Teller, *J. Amer. Chem. Soc.* **60**, 309 (1938).
18. National French Test Method, T 45-007 (1987).
19. S. J. Gregg and W. S. W. Sing, in *Adsorption, Surface Area and Porosity*, 2nd ed., Academic Press, London, U.K., 1982.
20. ASTM D3849, *Primary Aggregate Dimensions from Electron Microscopic Image Analysis*, ASTM, Philadelphia, Pa., 1980.
21. D. B. Williams, *Practical Analytical Electron Microscopy in Materials Science*, Electron Optics Publishing Group, Mahwah, N.J., 1987.
22. G. Hulett, in J. Alexander, ed., *Colloid Chemistry*, Chemical Catalog Co., New York, 1926.
23. J. V. Walther and H. C. Helgeson, *Amer. J. Sci.* **277**, 1315 (1977).
24. C. M. Jephcott and J. H. Johnston, *Arch. Ind. Hyg. Occup. Med.* **1**, 323 (1950).
25. J. J. Denny, W. D. Robeson, and D. A. Irwin, *Can. Med. Assoc. J.* **40**, 213 (1939).
26. S. G. DeBussetti, M. Tschapek, and A. K. Helmy, *J. Electroanalyt. Chem.* **36**, 507 (1972).
27. R. K. Iler, in E. Matijevic, ed., *Surface and Colloid Science*, Vol. 6, John Wiley & Sons, New York, 1973.

28. Z. Z. Vysotskiĭ and co-workers, in D. N. Strazhesko, ed., *Adsorption and Adsorbents*, No. 1, John Wiley & Sons, New York, 1973.

29. R. K. Iler, *J. Colloid Interface Sci.* **75**, 138 (1980).

30. **U.S. Pat. 3,012,973** (Dec 12, 1961), R. C. Atkins (to E. I. du Pont de Nemours & Co., Inc.).

31. **U.S. Pat. 3,440,176** (Apr. 22, 1969), R. Sippel (to E. I. du Pont de Nemours & Co., Inc.).

32. **U.S. Pat. 2,574,902** (Nov. 13, 1951), M. F. Bechtold and O. E. Snyder (to E. I. du Pont de Nemours & Co., Inc.).

33. D. H. Napper and R. J. Hunter, *Med. Technol. Publ. Int. Rev. Sci. London*, **7**, 241 (1971).

34. J. A. Hockey, *Chem. Ind. London*, **2**, 57 (1965).

35. L. C. F. Blackman and R. Harrop, *J. Appl. Chem.* **18**, 37 (1968).

36. A. Snoeyink and W. Weber, in J. F. Danielli, M. D. Rosenberg, and D. A. Cadenhead, eds., *Progress in Surface and Membrane Science*, Vol. 5, Academic Press, Inc., New York, 1972.

37. D. Barby, in G. D. Parfitt and K. S. W. Sing, eds., *Characterization of Powder Surfaces*, Academic Press, Inc., New York, 1976.

38. *Current Industrial Reports*, Series MA28, U.S. Dept. of Commerce, Washington, D.C., 1994.

39. R. K. Iler, *Chem. Australia*, **53**, 355 (1986).

40. L. L. Hench and W. Vasconcelos, *Annu. Rev. Mater. Sci.* **20**, 269 (1990).

41. C. Okkerse, in B. G. Linsen, ed., *Physical and Chemical Aspects of Adsorbents and Catalysts*, Academic Press, Inc., New York, 1970.

42. S. Kondo, *Hyomen*, **10**, 321 (1972).

43. I. E. Neimark and R. Y. Sheinfain, *Silica Gel: Preparation, Properties and Uses*, Nauk Dumka, Kiev, CIS, 1973.

44. H. Teicher, in T. C. Patton, ed., *Pigment Handbook*, Vol. 1, John Wiley & Sons, New York, 1973.

45. C. J. Brinker and co-workers, *J. Non-Cryst. Solids*, **48**, 47 (1982).

46. M. W. Colby, A. Osaka, and J. D. MacKenzie, *J. Non-Cryst. Solids*, **99**, 129 (1988).

47. C. J. Brinker and G. W. Scherer, *Sol-Gel Science: The Physics and Chemistry of Sol-Gel Processing*, Academic Press, Inc., Boston, 1990.

48. T. W. Zerda, I. Artaki, and J. Jonas, *J. Non-Cryst. Solids*, **81**, 365 (1986).

49. L. W. Ketls, N. J. Effinger, and S. M. Melpolder, *J. Non-Cryst. Solids*, **83**, 353 (1986).

50. L. Kutik, *J. Coat. Technol.* **58**, 91 (1986).

51. C. J. Brinker and co-workers, *Mater. Res. Soc. Symp. Proc.* **284**, 469 (1993).

52. J. Fricke and A. Emmerling, *Adv. Materials*, **3**, 504 (1991).

53. L. L. Hench, *Ceram. Trans.* **29**, 591 (1993).

54. P. Kleinschmit, *Spec. Inorg. Chem.* **40**, 196 (1981).

55. A. W. Kiselew, *J. Colloid Interface Sci.* **28**, 430 (1968).

56. H. Ferch, *Chem. Ing. Technol.* **48**, 922 (1976).

57. G. Taylor and J. H. Adkins, *J. Phys. Chem.* **70**, 1678 (1989).

58. E. P. Plueddermann, *Silane Coupling Agents*, Plenum Press, New York, 1982.

59. S. J. Monte and G. Sugerman, *Proceedings from the 33rd Technology Conference*, Reinforced Plastics Composite Institute, Washington, D.C., 1978.

60. W. H. Waddell and co-workers, *J. Appl. Polymer Sci.* **55**, 1627 (1995).

61. J. H. O'Haver and co-workers, *J. Appl. Polymer Sci.* **59**, 1427 (1996).

62. B. M. Cooper, *Ind. Miner.* **258**, 43 (Mar. 1989).

63. M. P. Wagner, *Rubber Chem. Technol.* **49**, 703 (1976).

64. R. F. Wolf and C. Stueber, *Rubber Age*, **87**, 1001 (1960).

65. L. R. Evans and W. H. Waddell, *Rubber Plast. News*, 15 (Apr. 25, 1994).

66. T. A. Okel and W. H. Waddell, *Rubber Chem. Technol.* **67**, 217 (1994).

67. W. H. Waddell and L. R. Evans, *Rubber Chem. Technol.* **69** (Aug. 1996).

68. **U.S. Pat. 5,227,425** (July 13, 1993), R. Rauline (to Michelin).

69. **Eur. Pat. 645423A** (Sept. 7, 1994), J.-C. J. Kiln and co-workers (to Goodyear).

70. **U.S. Pat. 5,094,829** (Mar. 10, 1992), T. G. Krivak, T. A. Okel, and M. P. Wagner (to PPG Industries).

71. W. H. Waddell, P. A. Beauregard, and L. R. Evans, *Tire Technol. Int.* **95**, 14 (1995).

72. T. A. Okel and W. H. Waddell, *Rubber Chem. Technol.* **68**, 59 (1995).

73. R. A. Crawford and G. E. Alderfer, *Neut./Alkal. Papermaking*, 125 (1990).

74. M. J. Weighall, *J. Power Sources*, **34**, 257 (1991).

75. J. L. Boyer, *Batteries Int.* 60, (Oct. 1994).

76. M. Pader, *Cosmetic Toiletries*, **103**, 84 (1988).

77. S. K. Wason and J. W. Maisel, *Plast. Compound.* **3**, 21 (May-June 1981).

78. FAO-WHO Joint Expert Committee on Food Additives, 17th Meeting, *World Health Org. Tech. Report Ser.* **539**, 16 (1974).

79. R. Oelmueller and H. Ferch, *ASTM Spec. Tech. Publ.* **1036**, 85 (1989).

80. R. D. Kulkarni, E. D. Goddard, and B. Kanner, *Ind. Eng. Chem. Fundam.* **16**, 472 (1977).

81. G. D. Ulrich, *Combust. Sci. Technol.* **4**, 47 (1971).

82. H. Cochrane and C. S. Lin, *Rubber World*, **192**, 29 (1985).

83. B. B. Boonstra, H. Cochrane, and E. M. Dannenberg, *Rubber Chem. Technol.* **48**, 558 (1975).

84. J. C. Saam, *Mat. Res. Soc. Symp. Proc.* **274**, 91 (1992).

85. D. G. Miller, W. F. Moll, and V. W. Taylor, *Mod. Paint Coat.* **73**, 53 (1983).

86. H. Cochrane and D. G. Miller, *Mod. Paint Coat.* **73**, 159 (1983).

87. *Cosmetic Chemicals: Worldwide Specialty Chemicals*, SRI International, Menlo Park, Calif., 1989.

88. D. Riffel, *Pigm. Resin Technol.* **15**, 13 (1986).

89. T. A. Bolton and G. A. Geineccius, *Perfumer Flavor.* **17**, 2 (Mar.-Apr. 1992).

90. V. V. Rondinella, M. J. Matthewson, and C. R. Kurkjian, *J. Amer. Ceram. Soc.* **77**, 73 (1994).

91. *Chemical Economics Handbook Marketing Research Report, Silicas and Silicates*, SRI International, Menlo Park, Calif., 1993.

92. H. E. Bargna, *Ind. Quim.* **28**, 123 (1970).

93. S. Mann and C. C. Perry, *Structural Aspects of Biogenic Silica*, Ciba Foundation Symposium 121, John Wiley & Sons, Inc., New York, 1986, pp. 40-58.

94. J. B. Jones and E. R. Sequit, *J. Geol. Soc. Australia*, **18**, 57 (1971).

95. R. V. Dietrich and B. J. Skinner, *Rocks and Rock Minerals*, John Wiley & Sons, New York, 1979.

96. R. O. Fournier, in *Proceedings of the Symposium on Hydrogeochemistry and Biogeochemistry*, Clarke Co., Washington, D.C., 1973.

97. R. H. Wilson to D. H. Groth, *ASTM Spec. Tech. Publ.* **732**, 143 (1981).

98. R. K. Iler, *ASTM Spec. Tech. Publ.* **732**, 9 (1981).

99. D. H. Groth and co-workers, *ASTM Spec. Tech. Publ.* **732**, 118 (1981).

100. G. W. H. Schepers and co-workers, *Arch. Ind. Health*, **16**, 125 (1957).

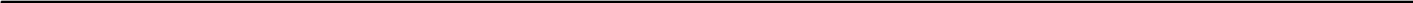
General References

R. K. Iler, *The Chemistry of Silica: Solubility, Polymerization, Colloid and Surface Properties, and Biochemistry*, John Wiley & Sons, New York, 1979.

P. Kleinschmit, *Special. Inorg. Chem.* **40**, 196–225 (1981).
H. D. Gesser and P. C. Goswami, *Chem. Rev.* **89**, 765–788 (1989).
L. L. Hench and W. Vasconcelos, *Ann. Rev. Mater. Sci.* **20**, 269–298 (1990).

Walter H. Waddell
Exxon Chemical Company

Larry R. Evans
J. M. Huber Corporation



VITREOUS SILICA

Vitreous silica [60676-86-0] is an amorphous phase of silicon dioxide, traditionally formed by heating crystalline SiO₂ above its melting point, 1730 ± 5° C (1), and then cooling it rapidly enough to avoid recrystallization. Because molten SiO₂ is extremely viscous, glass formation is possible at a relatively slow cooling rate. Vitreous silica is available in a variety of forms, including powders, coatings, fibers, porous bodies, and bulk (dense) glass. This article focuses, for the most part, on the processing, properties, and uses of the bulk glass (see also SILICA, AMORPHOUS SILICA).

Vitreous silica has a wide range of commercial and scientific applications. Its unique combination of physical properties includes good chemical resistance, minimal thermal expansion, high refractoriness, and excellent optical transmission from the ultraviolet to the near-infrared.

Although vitreous silica is a simple, single-component glass, its properties can vary significantly, depending on thermal history, the type and concentration of defects, and impurities. Vitreous silica can, however, be one of the purest commercially available glassy materials. In synthetic vitreous silicas, for example, total metal contamination is typically measured in the 50–100 ppb range. Even at such a low level of impurities, differences in properties, such as uv-transmission, are observed for various silicas.

Vitreous silica is a difficult material to manufacture, and as such is more expensive than most commercial glasses. It has a high vapor pressure, a strong tendency to devitrify, and a high viscosity even at its melting point. These physical characteristics preclude the use of standard glass-forming techniques. In combination with the need to maintain high purity and the continuing demand to improve the homogeneity of the refractive index, these characteristics have led to the development of several unique manufacturing approaches, such as electric fusion and flame hydrolysis.

Vitreous silica is, for the most part, a synthetic material, but there are instances of the material occurring in nature (2,3). Vitreous tubes called fulgurites are produced when lightning fuses quartz sand. Large deposits of fulgurite exist in the Libyan desert. Vitreous silica can also be produced by meteor impact. The impact leads to rapid adiabatic heating of the quartz above its melting point. The quartz forms a glass on cooling. Examples of this type of vitreous silica have been found near Canyon Diablo, Arizona, and in meteorite craters in Australia and Arabia.

Vitreous silica is known by a number of names, most of which were developed to describe vitreous silicas formed using a particular processing method (4–6). These names are often misapplied as generic names for all vitreous silicas. Silica (qv) is a general term which refers to all forms of silicon dioxide, both crystalline and glassy; silica glass is an ambiguous term because it normally refers to any silica-containing glass composition; fused silica applies to vitreous silicas formed by fusion; fused quartz refers to vitreous silica formed specifically by the direct melting of quartz crystals; and synthetic fused silica describes vitreous silicas formed from chemical precursors, such as silicon tetrachloride. Uses of the terms quartz and quartz glass should normally be avoided. These confuse vitreous silica with quartz, a crystalline form of silicon dioxide.

The optical quality of vitreous silica can range from transparent to opaque. The degree of transparency depends on the amount of bubbles present (7). Opaque glass contains a large number of isolated bubbles ranging in size from 5 to 200 µm. The bubbles scatter light, giving the glass a milky appearance. This material, produced by melting quartz sand, tends to be less pure (99.5–99.9 wt % SiO₂) than the transparent material. The opaque glass is an inexpensive form of vitreous silica and is often used where purity and optical properties are not important.

Transparent vitreous silicas are normally classified by an informal scheme which differentiates the glasses by manufacturing method (8–10). Five general types of commercial glasses are listed in Table 1. Although this list gives a good overview of the available glasses, it is not complete. Many of the listed manufacturing methods have processing variations which can alter glass chemistry and properties. Moreover, the Type V designation is not as widely recognized as the others. Also, the list does not include the hybrid processes which combine some of the key features of different methods. For example, there have been efforts to produce glass by flame-fusing sol-gel produced powder (11).

Table 1. Classification of Commercial Vitreous Silicas

Type	Method	Impurities, ppm			Uv cutoff ^a , nm	Representative glasses
		OH	Cl	Cations		
I	electric melting of quartz	<20	0	50–300	220	Infrasil, Vitreosil-ir, GE-124, GE-214
II	flame fusion of quartz	200–500	0	10–50	210	Homosil, Optosil, Vitreosil-O55
III	flame hydrolysis	600–1200	50–100	<1	170	Corning 7940, Dynasil 1000, Shinetsu P-10, Spectrosil, Suprasil, NSG-ES
IV	oxidation of SiCl ₄ (plasma)	<20	<200	1–2	170	Spectrosil WF, Suprasil-W
V	sol-gel	<1	<1–500	<1	170	GELSIL

^a Wavelength at which transmission drops to 50%.

Structure

Vitreous silica is considered the model glass-forming material and as a result has been the subject of a large number of x-ray, neutron, and electron diffraction studies (12–16). These investigations provide a detailed picture of the short-range structure in vitreous silica, but questions about the longer-range structure remain.

The basic structural element in both vitreous and crystalline silica is the SiO⁴⁺₄ tetrahedron, which arises from the *sp*³ hybrid orbitals of the silicon. Each silicon atom sits in the center of the tetrahedron surrounded by four oxygen atoms that hold the corner positions. Tetrahedrons bond together by corner sharing. In a properly developed structure, each oxygen is shared by only two tetrahedrons. This bonding scheme can produce a large variety of three-dimensional structures and is the reason that silica has a number of crystalline phases.

The bonding scheme can also accommodate a large degree of disorder without breaking the Si–O links. Diffraction studies on vitreous silica show that the randomness of the structure results from variations in the Si–O–Si bond angle. The individual SiO⁴⁺₄ units do not show significant distortion. The O–Si–O bond angle is fixed at 109.5°, expected in a symmetric tetrahedral configuration, and the Si–O bond distance, 161 pm, is the same as that observed in the crystalline silica phases. The Si–O–Si bond angles, on the other hand, show a distribution that varies from 120 to 180°. Whereas original analysis of the x-ray measurements placed the maximum of the angle distribution at approximately 144°, a more recent reexamination of the data showed

the most probable bond angle to be 152° (17). High temperature x-ray measurements show that the bond distances are constant up to 1000°C (18). In the crystalline form of silica, the SiO⁺₄ tetrahedrons link together in a series of interconnecting, six-membered rings. In the glassy form, the tetrahedral network can be modeled using a distribution of ring sizes normally ranging from three- to eight-membered units (19). The six-membered ring is the most common size but it can exist in distorted configurations which are not present in the crystalline structures. Detailed ring statistics of the glassy phase cannot be extracted directly from diffraction data. The three- and four-membered rings have been studied using Raman scattering (20,21). The longer-range structure of vitreous silica has been the subject of much debate. Two basic models have been presented (14,22,23). The random network model assumes that the SiO⁺₄ tetrahedrons are connected in a completely random fashion and that there are no ordered microregions. The crystallite-based models argue that vitreous silica consists of very small, randomly oriented microcrystals. These regions are assumed to be 0.7–2.0 nm in size and connected by noncrystalline boundary regions. One problem of the crystallite models is that, because of the small sizes proposed for the crystallites, the fraction of atoms in the boundary regions should be significant and include species such as O_ς=Si double bond, where O_ς signifies an oxygen on the surface of the crystallite. There are, however, no indications of large concentrations of these species in the diffraction and vibrational spectra as formed vitreous silica (24,25). As a counterpoint, several researchers have suggested that the anomalous changes seen in a number of the physical properties of vitreous silica near the transformation temperatures of the crystalline phases occur because the atomic arrangements present in vitreous silica relate, to some extent, to crystalline structures (26).

The tetrahedral network can be considered the idealized structure of vitreous silica. Disorder is present but the basic bonding scheme is still intact. An additional level of disorder occurs because the atomic arrangement can deviate from the fully bonded, stoichiometric form through the introduction of intrinsic (structural) defects and impurities. These perturbations in the structure have significant effects on many of the physical properties. A key concern is whether any of these defects breaks the Si–O bonds that hold the tetrahedral network together. Fracturing these links produces a less viscous structure which can respond more readily to thermal and mechanical changes.

The intrinsic defects include paramagnetic and diamagnetic species (24,27,28). The paramagnetic defects have received the most study because they are readily detectable by electron spin resonance (esr) spectrometry. Paramagnetic defects that have been identified by esr include the E' center,≡Si; the nonbridging oxygen hole center,≡Si–O; and the peroxy radical,≡Si–O–O, where the signifies an unpaired spin and≡represents the network bonding to the adjacent oxygens. All of these defects are normally induced in vitreous silica by ionizing radiation or uv light.

A number of diamagnetic defects are also believed to exist in vitreous silica. Because there is no direct way to study these species, their identification is either done indirectly, such as by uv absorption, or by employing esr after the material has been made paramagnetic using ionizing or laser irradiation. The proposed diamagnetic species include the neutral oxygen vacancy,≡Si–Si≡, the doubly coordinated silicon, –O–Si–O–; and the peroxy bond, ≡Si–O–O–Si≡. The diamagnetic defects occur when the glass-forming conditions are off-stoichiometry.

The impurity (extrinsic) defects include nonmetallic and metallic contaminants (29–32). The nonmetallic contaminants are the more prevalent type in most high grade vitreous silicas. In the synthetic silicas, for example, the concentrations of nonmetallics can be three to five orders of magnitude higher than that of metallics. The nonmetallic contaminants include species such as hydrogen, chlorine, fluorine, and carbon, which can be present in the furnace atmosphere during certain forming, consolidation, or special annealing treatments. These impurities exist in the vitreous silica in a variety of forms. They can link directly to the silica network, for example, hydroxyl,≡Si–OH, bonding; hydride,≡Si–H, bonding; and chloride,≡Si–Cl, bonding; or exist interstitially as molecular species, such as H₂, H₂O, Cl₂, and CO. Metallic contaminants are usually picked up from the silica raw material or from the furnace materials during processing. These dissolve in the silica network by bonding in some way to the oxygen ions. Table 2 lists the metal impurities commonly seen in vitreous silicas.

Table 2. Impurity Concentrations in Vitreous Silicas^a

Impurity, ppm	Translucent vitreous silica	Transparent vitreous silica ^b		
		Type I	Type II	Type III ^c
aluminum	18–176	16	10–50	<0.05
antimony	<0.06	0.3	0.15	<0.005
arsenic	<0.01	<0.4	0.08	<0.005
boron	<0.1	<0.1	0.1	0–0.01
calcium	2.0	0.6	0.8–3.0	<0.1
chromium	<0.09	<0.01	1–2.0	<0.03
cobalt		0.05		0.01
copper	<0.15	<0.1	0.7	0.004
gallium			0–0.008	<0.02
gold	<0.03		0.0003	
hydroxyl	<200	<5	~180	<1250
iron	1	0.3	0.8	<0.12
lithium	1.3	1.0	0–2.0	<0.001
magnesium	0.1	0.1	0.2	<0.05
manganese	0.1	0.1	0.01	<0.01
mercury				<0.01
phosphorus	0.1–1.5	1.5	0.1	<0.001
potassium	1.0	0.7	0.8	<0.1
sodium	2.0–3.0	1.0	1.0	<0.1
titanium	1.1–63	1.1	0.8	<0.04
uranium	<0.04		0.0003	<0.001
zinc				<0.03
zirconium	1.0–1.5	1.5	0–0.1	<0.03

^a Data taken from manufacturers' publications.
^b See Table 1 classifications.
^c The metallic impurities of Type IV vitreous silica are similar to those of Type III material, except for a negligible hydroxyl content.

Manufacturing

Vitreous silica, a hard glass having a limited working range, cannot be produced using conventional glass-melting techniques. Rather, its high melting point and high melt viscosity have forced the development of a number of unique forming methods which utilize sintering or some type of deposition process. Many of the key physical properties depend on the specific forming process used because the forming process defines the material's thermal history and impurity profile.

History. The first reported instances of fusing silicon dioxide occurred early in the nineteenth century. In 1813, fusing of small crystals of quartz by injecting oxygen into an alcohol flame was described. This was followed in 1821 by the description of the fusing of quartz crystals using a hydrogen–oxygen torch (33). Properties of vitreous silica were first measured in 1839 (34). A glass produced using a hydrogen–oxygen flame exhibited remarkable strength, good elasticity, and excellent thermal shock resistance. This glass also had a tendency to volatilize and to devitrify. Later, the glass was shaped into tubes and bulbs. In 1887, a flame was used to draw fibers of vitreous silica (35).

Several efforts to develop commercial processes began around 1900 and were centered mainly in England, France, and Germany (35). It was during this early commercialization of vitreous silica that the distinction between the opaque and transparent glasses arose. The opaque material, called fused silica, was made by fusing sand. Resistance heating with carbon electrodes was an effective method for generating the desired temperature. The transparent material, called fused quartz, was made by fusing clear, selected quartz crystals. Achieving transparency by this approach required overcoming the problem of bubble formation in the glass. Crystalline quartz undergoes a transformation, ie, an inversion from its low to high crystalline form, at 575°C, which also includes a rapid volume change. If heating is not uniform, the crystals can fracture (splinter) and lead to air encapsulation and the formation of bubble lines. This problem was overcome in 1900 by using the fracturing phenomenon to produce a fine crystalline frit which could be subsequently flame-fused without splintering (36,37). Fracturing was promoted by heating the quartz crystals to a red heat and then quenching in water. An approach in which a glass boule was produced by injecting this powder into a flame was also advanced (38).

The production of vitreous silica from chemical precursors was first described in patents filed in 1934, including a fabrication method in which fine, high purity powders were produced by decomposing silanes (39). Forms were then cast from aqueous slips. More importantly, a flame hydrolysis process which used SiCl₄ as the chemical precursor was described (40). This latter approach led to a marked improvement in glass purity and served as the basis for the processes used in the 1990s to make synthetic vitreous silica.

Modern Manufacturing Techniques. Manufacturing techniques for making bulk vitreous silica are for the most part improved variations of the historical processes. The main exception is the sol–gel process (see SOL–GEL TECHNOLOGY). All processes involve the fusion or viscous sintering of silica particles. The particles can be in the form of a loose powder or a porous preform. The powders can be made from natural quartz or from the decomposition of chemical precursors, such as silicon tetrachloride, and tetraethylorthosilicate (TEOS). In some approaches, such as flame hydrolysis, the powder is produced and fused in a single step. The improvements made to these techniques deal mainly with the procedures used to prepare the powders, that is, to control purity and particle size, and the specific conditions under which the powders are consolidated.

Translucent Vitreous Silica. Translucent vitreous silica is produced by fusion of high purity quartz sand crystals (41,42). Sand is packed around a graphite rod through which a current is passed. The resistance heating produces a plastic mass which can be blown into molds, drawn into tubing, or shaped by rolling or pressing. Separation from the graphite rod is facilitated by gaseous products formed by interfacial reaction. Because the outside is sandy, the product is known as sand-surface ware. A matte finish is obtained by mechanical buffing. A glazed surface is produced by fusing the outside surface with an electric carbon arc or flame.

The Rotosil process employed by Heraeus and Heraeus-Amersil is used for the production of tubular or cylindrical shapes. It permits greater uniformity and dimensional control than the previous process. The quartz is washed in hydrofluoric acid and distilled water to remove impurities which can promote crystallization during manufacture. The sand is then placed in a rotating horizontal steel tube where it is held at the circumference by centrifugal force. A carbon arc is passed slowly down the center of the drum to fuse the sand. A modification, using open rotating molds on a vertical axis, allows formation of crucibles, beakers, and bowls.

For refractory applications, sand of moderate purity is fused by an electric arc into an ingot. The ingot is then crushed and ground to a size suitable for injection-molding or slurry-casting. A variety of shapes can be produced, including bricks, nozzles, and pouring tubes. After shaping, the pieces are dried slowly and then fired at 1150–1250°C for one to four hours. The materials have a white, opaque appearance.

Transparent Vitreous Silica. Clear, transparent, bubble-free vitreous silica may be obtained by melting natural quartz minerals, ie, fused quartz, by flame or plasma vapor deposition (synthetic fused silicas), and by sol–gel processing.

Fused Quartz. The fused quartz glasses correspond to the Type I and Type II vitreous silicas described in Table I. Type I silica is made by the direct melting of quartz powder using the resistance, ie, Osram process (43), or induction-heated furnaces (44). The powder is fed into the top of a tubular furnace and melted in either a molybdenum or carbon crucible surrounded by a protective inert or reducing gas. Tubing or rod can be drawn from the bottom of the crucible. Vacuum is sometimes used prior to fusion to minimize the gas content in the pore spaces and hot pressing is sometimes applied after fusion to reduce the remaining bubbles. Type II silica is formed using a process originally developed by Heraeus. Quartz powder is fed through an oxy-hydrogen flame and collected as a dense glass on a rotating fused-quartz tube (45). The glass is withdrawn slowly from the flame as the fused quartz builds up. Chlorine can be introduced during laydown of the glass to improve glass purity. The chlorine removes metal contaminants, such as aluminum, iron, copper, zinc, titanium, and the alkali and alkali-earth metals, by forming volatile chlorides (46).

The crystalline quartz powder is treated before use to improve purity and to eliminate the formation of gaseous inclusions during fusion. The powders are first washed in mixed acids, including hydrofluoric acid, to remove surface contaminants and then heated to at least 800°C and plunged into distilled water to promote extensive microcracking. The resulting powder is well suited for vacuum melting because it melts faster and expels gaseous impurities more completely than untreated quartz crystals.

Brazil continues to be an important source of quartz powder for fused quartz manufacturers. Namibia and the Malagasy Republic are also suppliers. In the United States, large deposits of crystalline quartz exist in West Virginia, Pennsylvania, and Missouri. Suppliers as of the mid-1990s included U.S. Silica and Unimin Corporation. Alternatives to the natural quartz powder, including fumed silica, flame-hydrolyzed soot, and sol–gel-derived powder, have also been tried as feedstock for these processes (11).

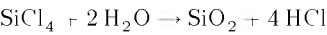
Although fused-quartz production is usually carried out above the melting point of cristobalite, powder sintering at temperatures in the range of 1400–1710°C is also possible (47). To avoid cristobalite formation (devitrification) while sintering at these temperatures, the powder must be pure, especially of alkali contaminants, and the time at temperature must be kept to a minimum. Typically, the sintering process takes no more than 15 min in the 1400–1710°C range. The sintering is usually done in a vacuum or in a helium atmosphere to minimize bubble formation.

Transparent vitreous silica crucibles can be produced by firing slip-cast pieces in a vacuum or a helium atmosphere (48,49), or in helium followed by argon (50). The crucible is usually supported on a graphite form to minimize distortion.

Synthetic Fused Silica. Synthetic silica can be prepared from a variety of volatile silicon compounds by oxidation or hydrolysis in a flame or plasma. Type III silica is made using a flame (40,51). In the United States, the flame is typically methane–oxygen. In Europe and Japan, hydrogen–oxygen flames are preferred. Type IV silica is made with a low water, oxygen plasma (52,53). The plasma deposition process was developed as a way to produce low OH vitreous silica. In glasses made by flame hydrolysis, the OH level ranges from 800–1200 ppm depending on the flame mix. In glasses made by plasma deposition, it is typically <20 ppm.

Vitreous silicas made from chemical precursors are significantly purer than those from natural quartz powders because it is possible to purify chemical precursors to a higher degree than powders. Synthetic fused silicas can have a total metal content below 200–300 ppb. Any stable, silicon-containing compound having sufficient volatility can be used as a chemical precursor (40,54). Most industrial processing uses silicon tetrachloride because it is relatively inexpensive and readily available in large quantities. Environmental concerns have, however, prompted the development of synthetic fused silica made from cyclic and chained siloxane compounds (55,56). These chlorine-free precursors do not generate HCl as a manufacturing by-product (see SILICON COMPOUNDS).

In the typical flame or vapor-phase hydrolysis process, which uses SiCl₄ as the precursor, silicon tetrachloride vapor is fed into a flame using an oxygen carrier gas. Fine (<50 nm), amorphous silica particles form in the flame (57).

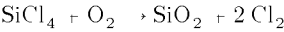


The water is a product of combustion of the hydrogen or methane gases.

The manufacture of vitreous silica by flame hydrolysis can be a single- or multistep process. In the single-step process, silica particles are formed and

consolidated in one operation. Consolidation takes place as the silica particles are collected on a substrate which is heated by the deposition flame. The deposition temperature is usually in excess of 1800°C. In a two-step process, the silica is formed as a soot which can be collected as a powder or porous preform (low temperature deposition) and consolidated in a separate operation. Consolidation is usually done at 1400–1700°C. The two-step process, although more involved, offers several advantages, including the ability to treat chemically the silica soot before final consolidation and to control the consolidation atmosphere. For example, high purity, low OH vitreous silica can be made by chemically drying the porous preform in chlorine (58) or SOCl₂ (59) and then consolidating in a moisture-free atmosphere. This operation is a key step in the production of optical waveguide fibers but can also be used to manufacture bulk glass (60,61).

In the plasma deposition process, silicon tetrachloride vapor is passed through an induction-coupled plasma torch using a hydrogen-free oxygen stream:



The resulting silica is deposited on a refractory substrate as a clear transparent glass.

Synthetic fused silicas with low OH levels have also been made experimentally using a CO₂ laser instead of the plasma torch (62,63). The glass rate of this process, however, is at least 10 times slower than the plasma approach.

Fused silica materials obtained by the various vapor-phase methods can be formed into rods, tubing, and boules. The pieces may range in size up to hundreds of kilograms. Any of these may be subsequently reworked to make a variety of useful products. Crucibles have been made from both tubing and massive pieces. Optical elements are cut from the boules and polished, and often finely annealed to control refractive index and reduce internal stress. *Sol–Gel Processing.* Sol–gel processing is basically a lower temperature alternative for fabricating synthetic fused silica (64–70). The starting material is either a particulate gel made by mixing fumed or colloidal silica in an appropriate liquid or a solution of a silicon alkoxide, such as tetraethylorthosilicate (TEOS) or silicon tetramethoxide, in alcohol using a controlled amount of water which hydrolyzes slowly to form a gel. After the gel is dried, it can be sintered under vacuum or in helium atmosphere to transparent vitreous silica at temperatures of 1200–1400°C. Steps are often taken to minimize the OH and chlorine content of the preforms to avoid bubble formation during sintering (71,72).

The gel approach offers processing advantages which are not available with the more conventional methods used to form synthetic fused silica. The lower firing temperatures reduce energy consumption and minimize the pickup of furnace impurities. Also, near-net-shaping is possible because the gelling procedure is not limited to any particular mold configuration. Unfortunately, the gel approach has critical drying and shrinkage problems which limit the glass sizes that can be produced. The gels, fragile structures that tend to crack during drying owing to capillary stresses, are highly porous structures that shrink linearly by up to 50% upon firing. The high shrinkage makes dimension control difficult. These problems have been attacked with limited success by increasing gel strength through gel aging, additions of silica particle fillers, etc; controlling the pore size in the gel through pH control and hydrothermal aging (73); and slower, milder drying procedures (see HYDROTHERMAL PROCESSING; SOL GEL TECHNOLOGY). In response to these processing limitations, several hybrid processes have been proposed that use sol–gel-derived powders as the feedstock for more conventional manufacturing approaches, including flame fusion and high temperature vacuum sintering (11,74).

Flame-Working and Sealing. Flame-working of vitreous silica is difficult because of its extremely high viscosity and volatility. However, this drawback is balanced by excellent resistance to thermal shock. Whereas a gas–oxygen flame is satisfactory for most manipulations, an oxy-hydrogen flame provides somewhat more energy. An oxy-acetylene flame gives even more heat, but promotes excessive volatilization.

When silica volatilizes, vapors condense on cooler areas to form a white bloom that can be removed by heat or dilute hydrofluoric acid. Because dilute hydrofluoric acid also attacks the substrate, a mild, careful treatment is required. To minimize volatilization, the temperature should be as low as possible.

Annealing of flame-worked pieces is generally not necessary because of the low thermal expansion. However, massive pieces should be annealed at a temperature corresponding approximately to the annealing point of the specific type and for a time determined by size. Owing to developments in applications of infrared lasers (qv), vitreous silica up to 6-mm thick can be cut and drilled using CO₂ gas lasers and numerically controlled equipment. Laser cutting has the advantage over mechanical cutting of a very thin kerf (cut width) having little induced stress (75).

Vitreous silica can be attached to other glasses, eg, Corning Code 7740 borosilicate glass, using a graded seal. The silica is first sealed to Code 7230 glass having a linear expansion coefficient, α , of ca 14×10^{-7} °C, from 0–300°C. This in turn is sealed to Code 7240 glass, $\alpha \approx 21 \times 10^{-7}$ °C. The assembly is then sealed to Code 7740 glass, $\alpha \approx 32.5 \times 10^{-7}$ °C.

Sealing to metals is difficult because the thermal expansion of metals is much higher than that of vitreous silica. Nevertheless, such sealing is necessary, particularly in the case of mercury vapor lamps where vacuum-tight lead connections are required. A molybdenum foil seal is used based on the principle that a very thin foil is able to contract during cooling without including stress in the glass. The molybdenum foil is connected to tungsten electrodes and placed in the tube which is flushed with a neutral gas. The tube is heated strongly on the outside with a gas–oxygen flame. It collapses around the foil and is mechanically pinched to make firm contact (76).

Glass Properties

Vitreous silica has many exceptional properties. Most are the expected result of vitreous silica being an extremely pure and strongly bonded glass. Inert to most common chemical agents, it has a high softening point, low thermal expansion, excellent thermal shock resistance, and an excellent optical transmission over a wide spectrum. Compared to other technical glasses, vitreous silica is one of the best thermal and electrical insulators and has one of the lowest indexes of refraction.

Vitreous silica, however, also exhibits a number of abnormal behaviors for a glass. These are a consequence of its inherent atomic structure (77–79). For example, the expansion coefficient is negative below approximately –80 to –100°C and positive and very small above these temperatures. In contrast to most glasses, the equilibrium density decreases with heat treatment in the transformation range. The elastic moduli increase with increasing temperatures above –190°C. The Young's modulus of vitreous silica increases linearly with applied longitudinal stress at –196°C, whereas the modulus of soda glass decreases (80). The compressibility of vitreous silica increases with pressure for pressures below 3 GPa (<30,000 atm) and the temperature coefficient of sound velocity is positive over the range 0–800°C (81).

Discontinuities in property–temperature relationships have also been observed and are often associated with structural rearrangements seen in the various crystalline phases of SiO₂ (26,82). The abnormalities in acoustic loss, pressure and temperature dependence of compressibility, thermal expansion, and specific heat have been related to a bimodal distribution of –Si–O–Si– bond angles (83). The large amount of free volume in the vitreous material permits reorientation of the silica tetrahedrons by changing –Si–O–Si– angles, ie, keeping Si–O distances constant, or by favoring the lower angles in the bimodal distribution.

Chemical Properties. Stoichiometric vitreous silica contains two atoms of oxygen for every one of silicon, but it is extremely doubtful if such a material really exists. In general, small amounts of impurities derived from the starting materials are present and various structural defects can be introduced, depending on the forming conditions. Water is incorporated into the glass structure as hydroxyls.

Chemical Durability. The resistance of nontransparent vitreous silica to chemical attack is slightly less than the resistance of transparent vitreous silica. This difference results primarily from the higher surface area of the former caused by the presence of a large number of bubbles. Most data in the literature are on the transparent material.

Vitreous silica does not react significantly with water under ambient conditions. The solution process involves the formation of monosilicic acid, Si(OH)₄. Solubility is fairly constant at low pH but increases rapidly when the pH exceeds 9 (84–86). Above a pH of 10.7 silica dissolves mainly as soluble silicates. Solubility also increases with higher temperatures and pressures. At 200–400°C and 1–30 MPa (<10–300 atm), for example, the solubility, *S*, of SiO₂ in g kg H₂O can be expressed as follows, where *d* is the density of the vapor phase and *T* is the absolute temperature in Kelvin.

$$\log S = 2 \log d - 2679/T + 4.972$$

Vitreous silica is susceptible to attack by alkaline solutions, especially at higher concentrations and temperatures. For 5% NaOH at 95°C, although crazing may be evident, surface corrosion is only 10 μm after 24 h (87). For 45 wt % NaOH at 200°C, dissolution proceeds at 0.54 mm/h (88). The corrosion rates in other alkaline solutions are listed in Table 3. Alkaline-earth ions inhibit alkaline solution attack on vitreous silica. Their presence leads to the formation of hydrated metal silicate films which protect the glass surface (90).

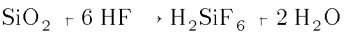
Table 3. Corrosion Rates of Vitreous Silica in Aqueous Media^a

Solution ^b	Temperature, °C	Duration of test, h	Weight loss, mg/cm ²
Alkaline solutions			
NH ₄ OH, 10%	20	100	0.019
NaOH			
10%	18	100	0.031
5%	95	6	0.7
8%	100	10	1.21
KOH			
30%	18	100	0.027
10.2%	100	10	1.13
Na ₂ CO ₃			
5%	18	100	0.0015
10%	100	10	0.37
Acid solutions			
acetic acid, 70%	108	24	0.1
hydrochloric acid			
5%	95	24	<0.01
ρ = 1.19	66	24	1.4
nitric acid, ρ = 1.40	115	24	1.1
sulfuric acid			
5%	95	24	<0.01
ρ = 1.84	205	24	0.6
phosphoric acid	300	15	580

^a Refs. 87 and 89.

^b ρ = density in g/mL.

Vitreous silica is relatively inert to attack from most acids for temperatures up to 100°C. The weight loss data in acid solutions are summarized in Table 3. The main exceptions are phosphoric acid, which causes some corrosion above approximately 150°C, and hydrofluoric acid, which reacts readily at room temperature (91). This latter dissolution proceeds as follows:



The equilibrium constant for this reaction at 25°C is 3.4 × 10⁻⁶ (92). The effects of two hydrofluoric acid solutions of different concentrations on various silica phases are shown in Table 4.

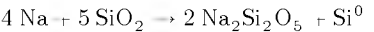
Table 4. Dissolution Rates of Silica Phases^a

Silica phase ^b	Silica dissolved, %	
	In 5% HF, 1–2 h	In 1% HF, 1 h
quartz	30.1	5.2
tridymite	76.3	20.3
cristobalite	74.3	25.8
vitreous silica	96.6	52.9

^a Ref. 93.

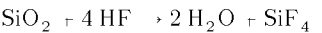
^b Samples of uniform particle size, ca 40 μm dia.

Metals do not generally react with vitreous silica below 1000°C or their melting point, whichever is lower. Exceptions are aluminum, magnesium, and alkali metals. Aluminum readily reduces silica at 700–800°C. Alkali metal vapors attack at temperatures as low as 200°C. Sodium vapor attack involves a diffusion of sodium into the glass, followed by a reduction of the silica.

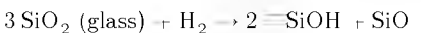
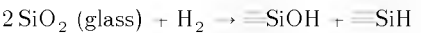


This reaction is accompanied by a blackening of the vitreous silica and a flaking of the surface. As evidence for the diffusion step, the sodium absorption by the much denser quartz is 12 mg/(1000 hcm²) at 350°C, whereas vitreous silica absorbs 23 mg/(1000 hcm²) at 286°C (94). Molten sodium is much less reactive.

Fused basic salts and basic oxides react with vitreous silica at elevated temperatures. Reaction with alkaline-earth oxides takes place at approximately 900°C. Halides tend to dissolve vitreous silica at high temperatures; fluorides are the most reactive (95). Dry halogen gases do not react with vitreous silica below 300°C. Hydrogen fluoride, however, readily attacks vitreous silica.

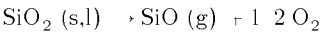


Reaction with hydrogen is very slight below 800°C, but reduction occurs at higher temperatures. In addition to some SiO formation, the formation of≡SiOH and≡SiH groups has been demonstrated by infrared and Raman spectroscopy (96).



Vapor Pressure. Vitreous silica is a highly refractory glass. Volatilization is significant only at elevated temperatures (Fig. 1), occurring under a

neutral atmosphere by dissociation.



SiO exists only as a vapor and reforms SiO₂ particles when it deposits on cold surfaces (97).

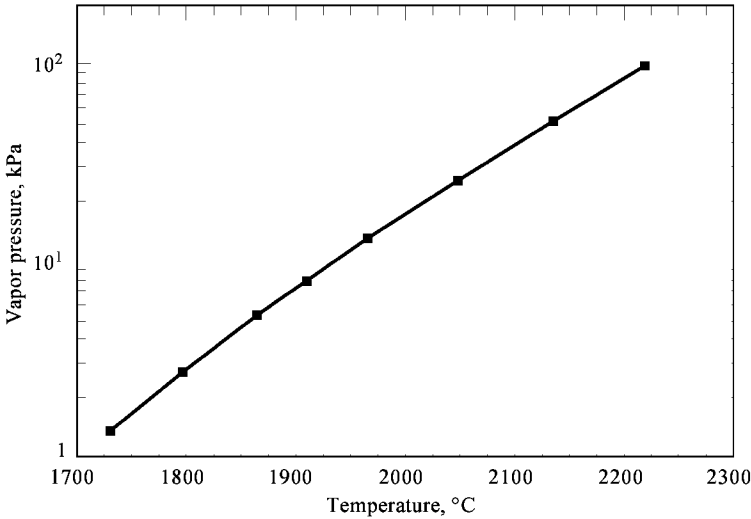


Fig. 1. Vapor pressure of silica at elevated temperatures (97). To convert kPa to psi, multiply by 0.145.

The loss of material at elevated temperatures is accelerated in vacuum and under reducing conditions. In the presence of carbon, for example, the following reaction may occur:



For pure carbon, this reduction can take place at temperatures as low as 1200°C (95). A similar reduction occurs with tungsten, tantalum, or molybdenum in vacuum at 1300–1400°C (98).

Devitrification. At atmospheric pressure, devitrification of vitreous silica can take place at temperatures from 1000 to 1723°C, ie, the cristobalite liquidus. The maximum growth rate occurs at approximately 1600–1675°C (Fig. 2). Crystals form and grow from nuclei found predominantly at the glass surface. Internal crystallization, though rarely seen, is also possible (1). Only heterogeneous nucleation has been observed as of this writing (1996), regardless of whether the crystallization occurs internally or externally. The crystalline phase, β-cristobalite, has nearly the same density and refractive index as vitreous silica. The β-to- α transition of cristobalite that occurs at 270°C includes a large volume change which leads to cracking and strength loss on cooling (100).

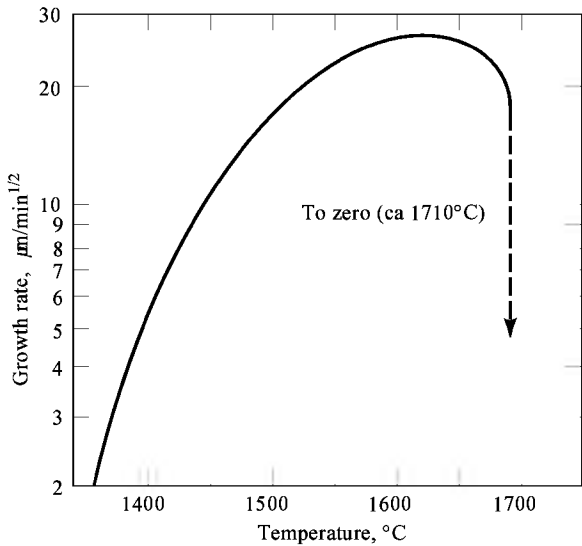


Fig. 2. Devitrification rate of vitreous silica for surface-nucleated cristobalite as a function of temperature (99). Growth rate is proportional to the square root of time. Measurements were made on G.E. fused silica, heat-treated in air.

The devitrification rate is extremely sensitive to both surface and bulk impurities, especially alkali. Increased alkali levels tend to increase the devitrification rate and lower the temperature at which the maximum rate occurs. For example, a bulk level of 0.32 wt % soda increases the maximum devitrification rate 20–30 times and lowers the temperature of maximum devitrification to approximately 1400°C (101). The impurity effect is present even at trace levels (<50 ppm) and can be enhanced with the addition of alumina. The devitrification rate varies inversely with the ratio of alumina-to-alkali metal oxide. The effect is a consequence of the fact that these impurities lower glass viscosity (102).

The water content and stoichiometry of the glass also influence the devitrification rate. A high hydroxyl content in the glass as well as water vapor and oxygen in the atmosphere enhance devitrification, whereas oxygen deficiency in the glass and neutral or reducing atmospheres inhibit devitrification (99,103–106). The oxygen affects the rate by diffusing through the cristobalite layer to the glass–crystal interface and oxidizing the glass, thereby bringing it closer to stoichiometry (103–105). The water vapor and hydroxyl content show a similar effect, probably by dissociating at the elevated temperatures to give free oxygen and by weakening the glass structure through the formation of silicon–hydroxyl bonds.

Increased pressures can lower the temperature at which crystallization occurs. Experiments performed using Spectrosil (Thermal Syndicate Ltd.) and G.E. Type 204 (General Electric Company) fused silicas (see Fig. 2) show that at pressures above 2.5 GPa (<25,000 atm), devitrification occurs at temperatures as low as 500°C and that at 4 GPa (<40,000 atm), it occurs at as low as 450°C (107). Although the temperatures and pressures were in the coesite-phase field, both coesite and quartz were observed. Both the devitrification rate and the formation of the stable phase (coesite) were enhanced by the presence of water. In the 1000–1700°C region at 500–4000 MPa (<5,000–40,000 atm), α- and β-quartz were the primary phases. Crystal growth rates

increased with pressure and with hydroxyl content for a given stoichiometry.

Cristobalite can also form on vitreous silica at temperatures as low as 400°C when the pressure is equal to 35 MPa (<350 atm) and the glass is immersed in weak NaOH solutions (108). In stronger NaOH solutions, quartz is formed. The formation of the crystalline phases is a result of the hydrolysis of the anions present. No crystallization occurs with HF, H₂SO₄, and H₃PO₄ in KHSO₄ solutions or in pure water.

Transparent silica can normally be used in air continuously at temperatures up to 1000°C and for short periods up to 1250°C without devitrification occurring. This recommendation assumes, however, that the glass surface is substantially free of alkali contamination which can occur from sources, such as airborne dust or fingerprints.

Diffusion. The gaseous species for which diffusion parameters have been studied extensively in vitreous silica include both noble gases and molecular gases. Data are given in Table 5. In general, the activation energies for diffusion of the gaseous species scale with molecular size (109). The noble gases do not interact chemically with the glass structure. Helium is the fastest diffusing species in vitreous silica. Its diffusion coefficient over the temperature range of 24–1034°C is as follows, where *T* is in Kelvin and *R* is the universal gas constant (26).

$$D_{\text{He}} = 2.7 \times 10^{-7} T \exp(-4.81/RT)$$

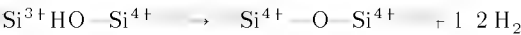
Table 5. Diffusion Coefficients in Vitreous Silica^a

Metal	Ion	<i>D</i> at 1000°C, cm ² s
lithium	Li ⁺	1 × 10 ⁻⁶
sodium	Na ⁺	7.9 × 10 ⁻⁶
potassium	K ⁺	(5–10) × 10 ⁻⁸
rubidium	Rb ⁺	(1.4–3.2) × 10 ⁻¹⁰
cesium	Cs ⁺	9 × 10 ⁻¹¹
calcium	Ca ²⁺	2 × 10 ⁻⁸
aluminum	Al ³⁺	1 × 10 ⁻¹³
nickel	Ni ²⁺	1 × 10 ⁻¹⁵

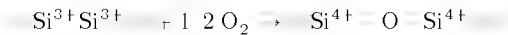
^a Refs. 112–114.

Diffusion of the molecular gases can be complicated by reactions with the glass network, especially at the sites of structural defects. The diffusion coefficient of water, for example, shows a distinct break around 550°C (110). Above 550°C, the activation energy is approximately 80 kJ mol⁻¹ (19 kcal mol⁻¹), but below 550°C, it is only 40 kJ mol⁻¹ (9.5 kcal mol⁻¹). Proposed explanations for the difference cite the fact that the reaction between water and the silica network to form hydroxyls is not in equilibrium at the lower temperatures.

The oxidation of vitreous silica appears to proceed by one of two mechanisms, depending on the material's hydroxyl content (109,111). In hydroxyl-containing material, the rapid oxidation probably occurs by the diffusion and removal of hydrogen, according to the following reaction:



Conversely, in hydroxyl-free vitreous silica, the oxidation is much slower and is controlled by the diffusion of oxygen through the solid according to the following reaction:



The diffusion of metal ions in vitreous silica has not been studied as extensively as that of the gaseous species. The alkali metals have received the most attention because their behavior is important in electrical applications. The diffusion coefficients for various metal ions are listed in Table 5. The general trend is for the diffusion coefficient to increase with larger ionic sizes and higher valences.

The diffusion of ²²Na in Infrasil fused quartz exhibits several distinct regions between 170 and 1000°C. The diffusion parameters are *D*₀ = 3.44 × 10⁻² cm² s, and *E*_a = 88.3 kJ mol⁻¹ (21.1 kcal mol⁻¹) between 573 and 1000°C; *D*₀ = 0.398 cm² s and *E*_a = 108 kJ mol⁻¹ (25.8 kcal mol⁻¹) between 250 and 573°C; and *D*₀ = 2.13 cm² s and *E*_a = 118 kJ mol⁻¹ (28.3 kcal mol⁻¹) between 170 and 250°C, where *D*₀ and *E*_a are the diffusion constant and activation energy of diffusion, respectively. The transitions occur at the transformation temperatures for the high–low modifications of quartz (573°C) and cristobalite (250°C), which suggests that the changes involve shifts in the specific volume of the glass structure (26,115). Diffusion data for gases in vitreous silica are listed in Table 6.

Table 6. Diffusion in Vitreous Silica^a

Gas	Diameter, pm	<i>D</i> at 1000°C, cm ² s	Activation energy, kJ mol ⁻¹ ^b
helium	200	5.5 × 10 ⁻⁵	20
neon	240	2.5 × 10 ⁻⁶	37
hydrogen	250	7.3 × 10 ⁻⁶	36
argon	320	1.2 × 10 ⁻⁹	113
oxygen	320	6.6 × 10 ⁻⁹	105
nitrogen	340		110
krypton	420		~190
xenon	490		~300
fluorine		3.3 × 10 ⁻¹⁶	383
chlorine		6.2 × 10 ⁻¹³	211

^a Refs. 116–118.

^b To convert J to cal, divide by 4.184.

Physical Properties.

Density. The density of transparent vitreous silica is approximately 2.20 g cm⁻³. Translucent and opaque glasses have lower densities owing to the entrapped bubbles. The density of translucent Vitreosil, for example, is 2.07–2.15 g cm⁻³ (87,119). The density of transparent vitreous silica decreases with increasing hydroxyl content and with lower fictive (glass structure equilibrium) temperatures. The fictive temperature depends on the thermal history and on glass viscosity (120).

The density of vitreous silica is lower than those of the low pressure crystalline phases of silicon dioxide (121), especially that of quartz.

Material	density, gm cm ³
----------	-----------------------------

vitreous silica	2.20
cristobalite	2.27
tridymite	2.30
quartz	2.65

Because all of these structures share the same short-range bonding scheme, the density differences indicate that vitreous silica has a substantial interstitial volume and can be compacted.

A number of studies have examined the high pressure behavior of vitreous silica (122–127). The degree of compaction depends on temperature, pressure, and the time of pressurization treatment. A density increase of almost 19% was obtained when the glass was subjected to a pressure of 8 GPa (<80,000 atm) at 575°C for 2 min. The samples remained completely amorphous, though the densities and refractive indexes were approaching those of the quartz phase (123,128). Structural studies indicate that the compaction is accompanied by a shift in the ring statistics toward a higher percentage of three- and four-membered rings of SiO₄ tetrahedrons (129). Above 20 GPa (200,000 atm), the SiO₄ tetrahedrons begin to distort and the Si coordination gradually shifts from four to six oxygens (130).

Compaction has also been observed as a result of neutron irradiation and extended exposure to intense uv (excimer) laser light (131,132). The compaction tends to relax over months at room temperature and can be reversed quickly by annealing at sufficiently high temperatures (133).

Viscosity. The viscosity of vitreous silica in the transformation range depends on impurities and thermal history (26,134,135). Any contaminant, including OH, chlorine, fluorine, and metal ions, that can break up the Si–O–Si network lowers the viscosity. Vitreous silicas with hydroxyl levels of approximately 0.1 wt %, such as Corning Code 7940, Suprasil, Spectrosil, and Dynasil, have annealing points in the 1050–1100°C range (119,136–138). Hydroxyl-free (<0.001 wt%) high purity fused-quartz glasses, such as infrared Vitreosil, some of the General Electric types, and various Amersil grades, have softening, annealing, and strain points up to 100°C higher than those of the hydroxyl-containing types. Thermal history determines the fictive temperature of the glass. Ir Vitreosil, which has a fictive temperature of 1400°C, has a viscosity of 1 TPas(10¹³ P) at 1700°C, whereas a 1000°C fictive material has a viscosity of 1 TPas(10¹³ P) at approximately 1320°C (134). Representative viscosity data are given in Table 7.

Table 7. Viscosity Data^a

Parameter	Quartz Syndicate, Inc.		G.E.-204	Vitreosil		Heraeus Supersil	Corning 7940	Dynasil
	Transparent	Translucent		ir	OG ^b			
silica type	I	I	II	II	I	III	III	III
OH, ppm			low	3	400	1200	900–1000	600–1000
softening point, °C	1670	1650	1813	158	159	1600	1585	1600 ± 25
annealing point, °C	1140	1100	1213	2	7	1075	1075	1100 ± 20
				119	110			
strain point, °C	1070	1040	1107	0	8	987	990	1000 ± 20
				110	101			
				8	5			

^a Data from manufacturers' product literature; data for Vitreosil from Ref. 134.
^b OG = optical grade.

A complete viscosity–temperature curve is shown in Figure 3 for Amersil commercial-grade fused quartz. A number of studies have shown that, above 1000°C, the viscosity, η , of vitreous silica follows the general relation

$$\ln \eta = \ln \eta_0 + E/RT$$

where η_0 is the viscosity at infinite temperature, T is temperature in Kelvin, E is the activation energy, and R , the molar gas constant, is 8.31 kJ/(mol·K) (141). Typical activation energy for viscous flow ranges roughly from 450 to 550 kJ/mol (108–132 kcal/mol) for the high OH vitreous silicas.

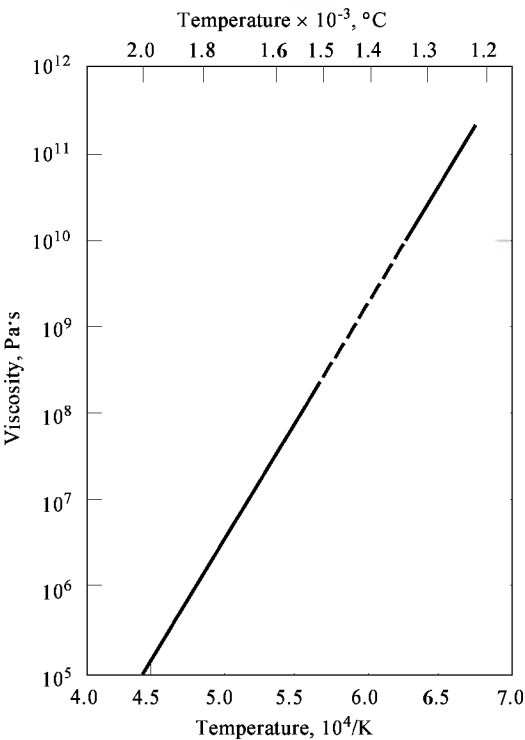


Fig. 3. Viscosity–temperature curve for Amersil commercial-grade silica where the dashed line represents interpolation between data sets (139,140). 10 Pas = 1poise.

The characteristic temperatures, ie, softening, anneal, and strain points, of vitreous silica are significantly higher than those of most commercial glasses. Vitreous silica starts to deform slowly at temperatures above 1200°C and, although it can be readily sagged at temperatures of 1400–1550°C, is a fairly viscous material even above the melting point of cristobalite (10⁶–10⁹ Pa·s (10¹⁰–10¹³ P)). The very high temperature viscosity data range from 930 Pa·s (9300 P) at 2085°C to 860 Pa·s (8600 P) at 3210°C (142).

Thermal Properties.

Thermal Expansion. Most manufacturers' literature (87,119,136–138) quotes a linear expansion coefficient within the 0–300°C range of 5.4 × 10^{−7} to 5.6 × 10^{−7} °C^{−1}. The effect of thermal history on low temperature expansion of Homosil (Heraeus Schott Quarzschmelze GmbH) and Osram's vitreous silicas is shown in Figure 4. The 1000, 1300, and 1720°C curves are for samples that were held at these temperatures until equilibrium density was achieved and then quenched in water. The effect of temperature on linear expansion of vitreous silica is compared with that of typical soda–lime and borosilicate glasses in Figure 5. The low thermal expansion of vitreous silica is the main reason that it has a high thermal shock resistance compared to other glasses.

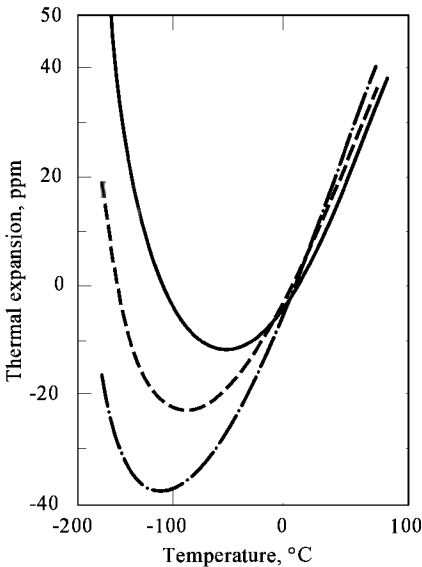


Fig. 4. Effect of thermal history on low temperature thermal expansion of vitreous silica (143), where (---), (— · —), and (—) represent glasses having fictive temperatures of 1000, 1300, and 1720°C, respectively.

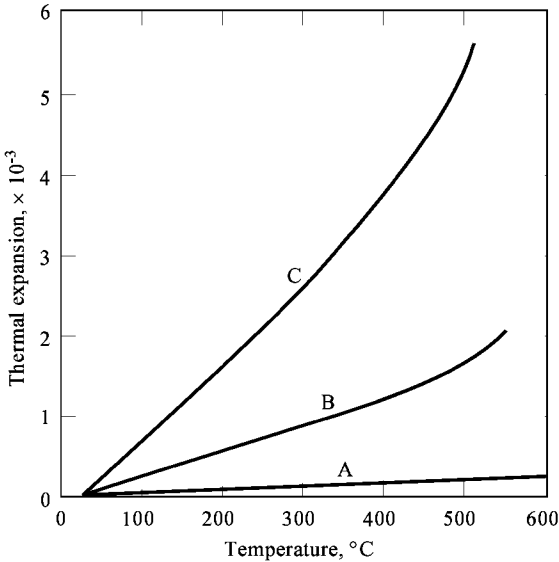


Fig. 5. Comparison of thermal expansion of vitreous silica, A, to that of borosilicate glass, B, and soda–lime glass, C.

Vitreous silica annealed at 1100°C has been designated NIST Standard Reference Material 739 (L1 and L2). Its expansion coefficient, α , may be calculated for 300–700 K from the following expression (144), where T is the absolute temperature in Kelvin.

$$\alpha \times 10^6 = 0.8218 + 7.606 \times 10^{-3} T - 1.266 \times 10^{-5} T^2 + 6.487 \times 10^{-9} T^3$$

Precise measurements of the dimensional stability of low expansion materials indicate that vitreous silicas, eg, Corning 7940 and Homosil, display a length change at 25°C of approximately 0.5 parts per billion (ppb) per day (145).

The expansion coefficient of vitreous silica can be controlled by doping the glass with titania. At 7.4 wt % TiO₂, the room temperature expansion coefficient is effectively zero (<10^{−8} °C^{−1}) (146).

Heat Capacity. The mean heat capacity (0–900°C) at constant pressure, C_p , in J (kg°C), can be estimated in vitreous silica using the following expression, where t is temperature in °C.

$$C_p = 700 + 0.79 t - 5.23 \times 10^{-4} t^2$$

Its value at 25°C is 0.71 J (g°C) (0.17 cal (g°C)) (95,147). Discontinuities in the temperature dependence of the heat capacity have been attributed to structural changes, eg, crystallization and annealing effects, in the glass. The heat capacity varies weakly with OH content. Increasing the OH level from 0.0003 to 0.12 wt % reduces the heat capacity by approximately 0.5% at 300 K and by 1.6% at 700 K (148). The low temperature (<10 K) heat capacities of vitreous silica tend to be higher than the values predicted by the Debye model (149).

Thermal Conductivity. Thermal conductivity data for transparent vitreous silica are listed below (150):

Temperature, K	Thermal conductivity, W (m·K)
100	0.674
200	1.126
300	1.37
400	1.51
500	1.62
600	1.74

Thermal conductivity at 25°C is 1.38 W (m·K). The thermal conductivity of opaque silica is 20% lower than that of clear vitreous silica.

Phonon transport is the main conduction mechanism below 300°C. Compositional effects are significant because the mean free phonon path is limited by the random glass structure. Estimates of the mean free phonon path in vitreous silica, made using elastic wave velocity, heat capacity, and thermal conductivity data, generate a value of 520 pm, which is on the order of the dimensions of the SiO₄ tetrahedron (151). Radiative conduction mechanisms can be significant at higher temperatures.

Thermal diffusivity, *a*, is related to the thermal conductivity, *k*, through the following equation, where ρ is density and *C_p* is heat capacity. The value of *a* at 25°C is 9 × 10⁻³ cm² s (152).

$$a = k / (\rho C_p)$$

Mechanical Properties.

Moduli and Poisson's Ratio. The Young's modulus of vitreous silica at 25°C is 73 GPa (<1.06 × 10⁷ psi), the shear modulus is 31 GPa (<4.5 × 10⁶ psi), and the Poisson's ratio is 0.17. Minor differences in values can arise owing to density variations. The elastic modulus decreases with increasing density and Poisson's ratio increases (26).

In contrast to the behavior of most glasses, the elastic moduli of vitreous silica increase with temperature, reaching a maximum at 1100–1200°C. The maximum is approximately 10% higher than the room temperature value (153). The high temperature values, however, are probably not accurate readings of the instantaneous moduli because viscous deformation is possible above 1000°C.

Strength. The fracture strength of vitreous silica depends on its surface quality, which can be affected by thermal treatment and handling conditions. Microcracks, surface contamination, and crystallization can reduce the strength from the value of pristine vitreous silica by several orders of magnitude.

The theoretical estimates of the intrinsic strength of vitreous silica, based on structural considerations, range from 1.6 GPa (<235,000 psi) to 2.3 GPa (<338,000 psi) (154,155). Experimental values close to theoretical numbers have been obtained under carefully controlled conditions. These conditions involve fibers having undamaged surfaces and measurements made in vacuum at low temperatures. For example, tests run at -196°C detected strengths in excess of 1.3 GPa (191,000 psi) (156,157). Conversely, the room temperature strength of a flame-drawn fiber is decreased from ca 4.7 GPa (<690 psi) to 344 MPa (<50,600 psi) by merely placing a finger on it (158).

The fracture strength of vitreous silica is also affected by environmental factors, such as temperature and humidity. The temperature effect on the strength of flame-drawn fiber is shown in Figure 6. The strength of fused silica rods decreases in the presence of saturated vapors of alcohols, benzene, acetone, and water (159). The largest decrease is caused by water vapor, ie, almost 50% from a starting strength of 91 MPa (<13,400 psi). Water vapor weakens vitreous silica by promoting crack growth. The critical fracture energy, *Y_c*, for example, is 4.3–4.4 J m⁻² (18–18.4 cal m⁻²) in dry nitrogen gas at 27°C, and 3.7 J m⁻² (15.5 cal m⁻²) in air at 40% rh (160,161).

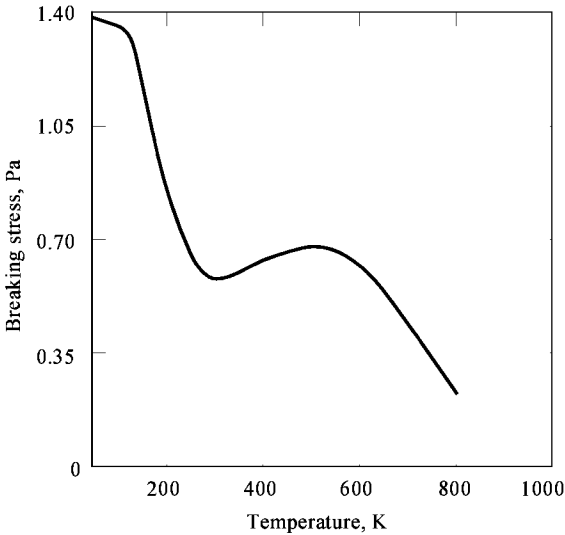


Fig. 6. Effect of temperature on the strength of vitreous silica fiber (156). To convert Pa to mm Hg, multiply by 0.0075.

Commercially available clear silicas typically have tensile strengths of 50–70 MPa (7,250–10,150 psi) and compressive strengths of 500–1900 MPa (72,500–275,500 psi). The opaque silicas have tensile strengths of 5–50 MPa (725–7250 psi) and compressive strengths of 190–300 MPa (27,550–43,500 psi) (162). Safety factors of 10–20-fold are usually employed when developing structural elements made of vitreous silica.

Fracture toughness is another way to characterize the strength of a material. It measures how well a material resists crack propagation and is expressed as the stress needed to enlarge a crack of a specific size. The room temperature fracture toughness of clear, vitreous silica is approximately 0.75–0.80 MPa·m^{1/2} (87,163).

Hardness. The Knoop indentation hardness of vitreous silica is in the range of 473–593 kg mm⁻² and the diamond pyramidal (Vickers) hardness is in the range of 600–750 kg mm⁻² (164). The Vickers hardness for fused quartz decreases with increasing temperature but suddenly decreases at approximately 70°C. In addition, a small positive discontinuity occurs at 570°C, which may result from a memory of quartz structure (165). A maximum at 570°C is attributed to the presence of small amounts of quartz microcrystals (166). Scanning electron microscopic (sem) examination of the indentation area indicates that deformation is mainly from material compaction. There is little evidence of shear flow (167).

Ultrasonic Properties. Vitreous silica of high purity, such as the synthetic type, has an unusually low attenuation of high frequency ultrasonic waves. The loss, *A*, is a linear function of frequency, *f*, up to the 30–40 MHz region and can be expressed as *A* = *Bf*, where *B* = 0.26 dB·MHz m for shear waves and 0.16 dB·MHz m for compressional waves (168).

The ultrasonic relaxation loss may involve a thermally activated structural relaxation associated with a shifting of bridging oxygen atoms between two equilibrium positions (169). The velocity, v , of ultrasonic waves in an infinite medium is given by the following equation, where M is the appropriate elastic modulus, and density, d , is 2.20 g cm³.

$$v = (M/d)^{1/2}$$

At a shear modulus of 31.1 GPa (4.5 × 10⁶ psi), vitreous silica has a plane-shear wave velocity of 3.76 × 10⁵ cm s.

Shear velocity and acoustic absorption have been studied as a function of OH content and fictive temperature for four different vitreous silica samples (170). All showed a shear-velocity minimum at 80–100 K. The magnitudes of the minima are influenced by OH content and fictive temperature. The samples having the highest OH content and lowest fictive temperature display the lowest losses.

Electrical Properties.

Conductivity. Pure vitreous silica is an excellent electrical insulator. The synthetic glasses can have bulk resistivities as high as 10¹⁸ ohm-cm at room temperature. Vitreous silica is an ionic conductor. Its large (8.1 eV) band gap precludes the formation of a significant concentration of electronic charge carriers (171). The main charge-carrying species are the monovalent cation impurities, especially sodium. Assuming an Arrhenius model, the activation energies of the conductivity at 350°C from sodium, lithium, and potassium ions are 85.9, 122.8, and 131.6 kJ mol (20.5, 29.3, and 31.5 kcal mol), respectively (172). The temperature dependence of the electrical conductivity, however, does not appear to follow strict Arrhenius behavior. The activation energy tends to decrease with increasing temperature. Part of the reason for this deviation may be glass-electrode effects (173).

The surface conductivity of vitreous silica is also low compared to other silicate glasses. Because vitreous silica is not hygroscopic, water films containing exuded alkalis do not readily form on its surfaces. The surface conductivity, however, can increase significantly with increasing relative humidity. A change in the relative humidity from 20 to 80% produces a millionfold increase in the surface conductivity (174).

Dielectric Properties. Vitreous silica is also a very stable dielectric material. The dielectric constant is 3.8–4.0, depending on the grade of glass. The loss factor, usually less than 0.00004, is the lowest available in glassy materials (87). The temperature and frequency dependences of both properties in a typical synthetic-grade vitreous silica are shown in Table 8. Lower purity silicas have slightly lower permittivities and higher loss factors. The loss factor is sensitive to impurities, especially the OH content of the glass (175). A discontinuity in loss factor occurs near 570°C, the quartz transition temperature. This may reflect the presence of quartz-like regions in the glass.

Table 8. Dielectric Properties of Synthetic Fused Silica^a

Parameter	Frequency, Hz		
	100	1,000	10,000
dielectric constant			
25°C	3.79	3.79	3.79
200°C	3.81	3.81	3.81
300°C	3.82	3.82	3.82
400°C	3.85	3.83	3.82
loss factor			
25°C	0.00002	0.00002	0.00002
200°C	0.00052	0.00012	0.00004
300°C	0.08	0.0072	0.00072
400°C	1.0	0.2	0.022

^a Ref. 87.

The dielectric breakdown strength in vitreous silica depends on its impurity content, its surface texture, and the concentration of structural defects, such as cord and bubbles. Good quality glasses have room temperature breakdown strength in the range of 200–400 kV cm.

Optical Properties. The optical transmission of vitreous silica is influenced by impurities and the forming process. Ultrapure vitreous silica has the ability to transmit from the deep ultraviolet, through the visible, and into the near-infrared spectral range.

The ultraviolet cutoff or the absorption edge for pure vitreous silica is 8.1 eV or 153 nm (171). This uv cutoff is influenced by the impurity level and stoichiometry of the material. Several impurities, such as the transition metals (Fe, Cu, Ti, etc) and alkali metal ions (Na, Li, K), degrade the ultraviolet performance, shifting the uv cutoff to longer wavelengths. Ferric ions (Fe³⁺) cause absorption or result in network defects under reducing conditions. This contaminant at only a few ppm can be detected as an absorption at 230 nm and below (176).

An absorption band at 242 nm is characteristic of reduced vitreous silica manufactured by fusing crystalline quartz powders. This band is apparent in Type I and II silica, such as infrared Vitreosil, G.E.-124 fused quartz, Infrasil, Homosil, Herasil (Optosil), and most of the NSG quartz. The 242-nm absorption has been attributed to a number of defects that include Si²⁺ (177), Si³⁺ associated with Al³⁺ (178), and germanium-related oxygen-deficient centers (179,180).

The Si–O oscillation produces strong absorption bands at 8.83, 8.98, 12.41, and 21.37 μm. The fundamental vibration at 8.83 μm also has a strong overtone at 4.45 μm, ie, the practical infrared cutoff wavelength. Water incorporated in the silica structure gives rise to a very strong absorption at 2.73 μm, which is observed in Type I and II glass. A combination of the SiO₄ fundamental at 12.41 μm and the strong OH fundamental at 2.73 μm produces a peak at 2.22 μm observed in the high water content silicas (181).

Vitreous silica transmission curves are shown in Figure 7. The hydroxyl concentration for each silica type is listed in Table 9. These curves represent only the general characteristics of the different silica types and should not be used for calculations of transmittance.

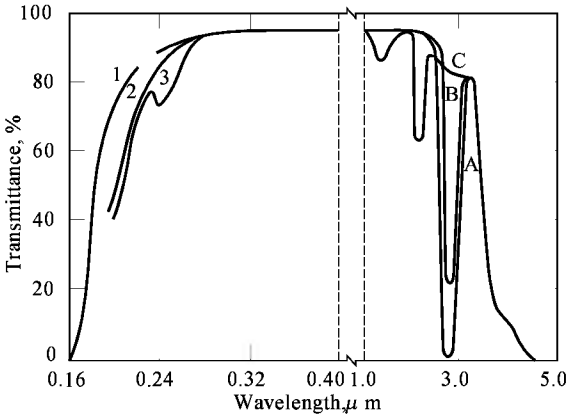


Fig. 7. Transmission curves for vitreous silica, 1-cm thick. See Table 9.

Table 9. Hydroxyl Concentration of Vitreous Silicas

Manufacturer	Material name	Uv curve ^a	Ir curve ^a	β _{OH} , ppm
Corning, Inc.	Code 7940	1	A	800–1000
	Code 7943	3	C	<1
	Code 7957	1	C	<1
Dynasil Corp. of America	Dynasil			
	Petrosil ^b	1	A	<1000
	QIR	3	C	<0.1
	QUV	2	B	150–180
	QVIS, QZ	3	B	150–180
	SWF	2	C	5
	SUV	1	A	1200–1500
General Electric Co.	Types 214 rods	2	C	<5
	Types 124 ingots	3	C	<10
Heraeus Amersil Inc.	Ultrasil	2	B	180
	Homosil, Optosil	3	B	180
	Infrasil	3	C	<8
	Suprasil-W1, W2	1	C	5
	Suprasil 1,2,3	1	A	900–1200
	Suprasil 300	1	C	<1
	Suprasil 311, 312	1	C	200
Nippon Silica Glass	ES	1	A	1200
	IR	3	C	<8
	SG, OX	3	B	150
	Spectrosil A, B	1	A	1000
	Spectrosil WF	1	C	
Quartz & Silice ^c (Thermal American Fused Quartz)	Vitreosil ir	3	C	
Shinetsu Quartz (JV with Heraeus)	Suprsil P10, 20, 30	1	A	1200
SumiQuartz (Sumitomo)	SK-1300	1	B	150
Westdeutsche Quarzschmelze GmbH	Synsil	1	A	

^a See Fig. 7.
^b Petrosil is manufactured in Russia.
^c Quartz & Silice, Quartz Products Co., and Thermal Quartz Schelze GmbH are Saint-Gobain Group Companies.

The spectral normal emissivity of Corning Code 7940 vitreous silica was computed from measurements of transmittance and reflectance at 25°C and is shown in Figure 8. The total normal emissivity of Code 7940 silica as a function of temperature from 200 to 1400°C is shown in Figure 9.

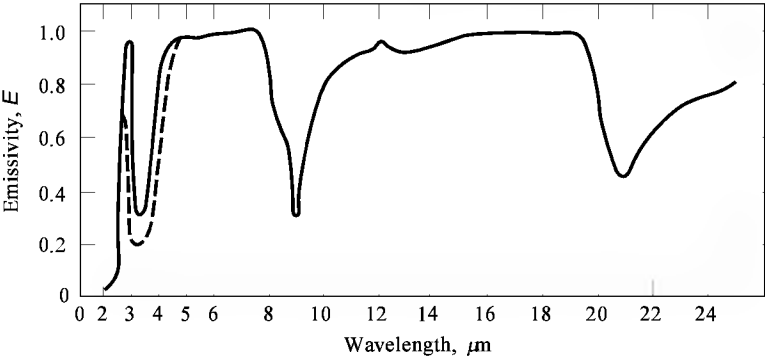


Fig. 8. Special normal emissivity of vitreous silica, where (—) corresponds to a sample 1.27-cm thick and (---), 0.64-cm thick (182).

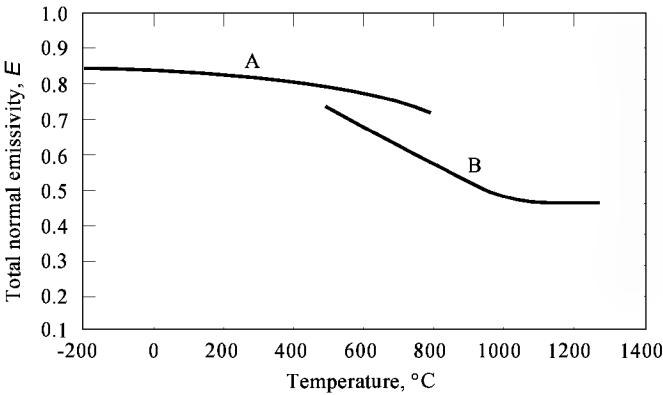


Fig. 9. Total normal emissivity of vitreous silica, where A corresponds to a sample 1.27-cm thick and B, 0.64-cm thick (183).

The index of refraction of synthetic vitreous silica at 20°C has been fitted to a three-term Sellmeier dispersion equation for wavelengths from 0.2139 to 3.7067 μm (184), where *n* is the refractive index and λ is the corresponding wavelength in micrometers.

$$n_D^2 - 1 = \frac{0.6961663 \lambda^2}{\lambda^2 - (0.0684043)^2} + \frac{0.4079426 \lambda^2}{\lambda^2 - (0.1162414)^2} + \frac{0.8974794 \lambda^2}{\lambda^2 - (9.896161)^2}$$

The computed refractive index at 20°C is 1.534307 at 0.213856 μm, 1.508398 at 0.248272 μm, 1.474539 at 0.365015 μm, 1.45637 at 0.65627 μm, and 1.39936 at 3.7067 μm. Using suspect data removed from the computed index, the predicted index fit with an average deviation of 4.3 × 10⁻⁶ over the wavelength range of 0.21–2.32 μm (185). The Abbe constant, which describes the reciprocal of the dispersive power, is 67.8. Refractive index also changes with temperature. The average values of Δ*n*_D/Δ*T* at 25°C are shown in Figure 10. The Δ*n*/Δ*T* values as a function of temperature from -200 to 40°C are shown in Figure 11 for vitreous silica at 587.6 nm. The stress optical coefficient, sometimes referred to as stress birefringence constant, is 3.45 nm/(cm·Pa). Vitreous silica refractive index is affected by the chlorine and hydroxyl concentration. An increase in the hydroxyl concentration of 1 ppm results in a decrease in refractive index of 1 × 10⁻⁷ (187). Another factor that can influence the refractive index is the thermal history of the glass (188,189).

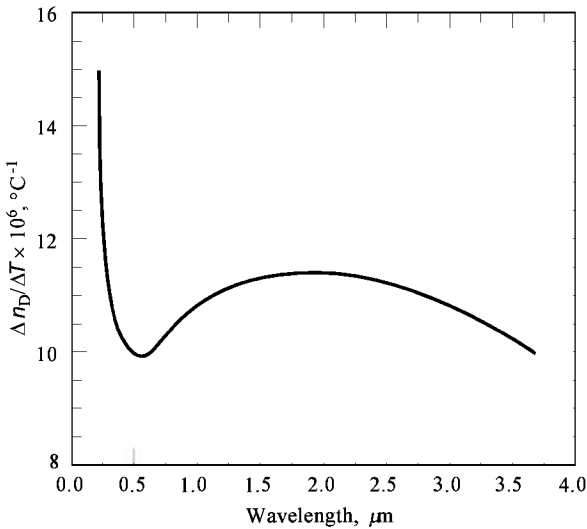


Fig. 10. Thermal coefficient of the refractive index in vitreous silica at approximately 25°C (184).

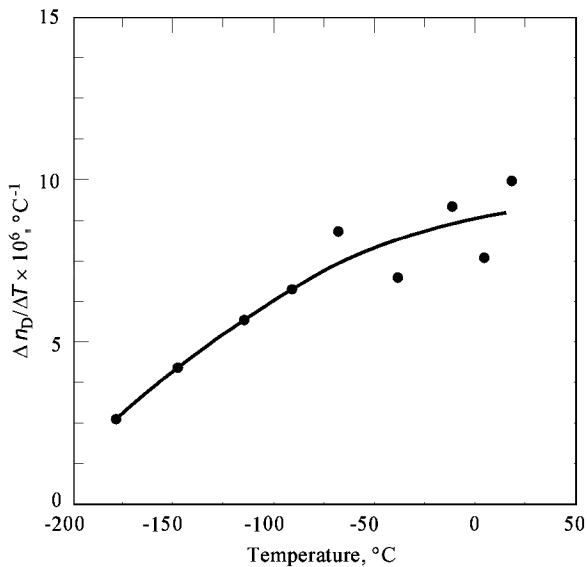


Fig. 11. Change in Δ*n*_D/Δ*T* with temperature in vitreous silica measured at 587.6 nm (186).

Radiation Effects. Depending on the type of radiation, significant structural changes are generally a result of high energy particle radiation, eg, a neutron stream, whereas electronic changes are usually the result of ionizing radiation, such as electron beams, x-rays, γ-rays, or protons. Damage by ionizing radiation manifests itself in the formation of optical absorption centers, also called color centers or defect centers. Although the predominant effect of neutron radiation is structural, electronic changes can also occur and, if the ionizing radiation is of very high intensity, structural defects may be produced.

Irradiation by fast neutrons causes a densification of vitreous silica that reaches a maximum value of 2.26 g/cm³, ie, an increase of approximately 3%, after a dose of 1 × 10²⁰ neutrons per square centimeter. Doses of up to 2 × 10²⁰ n/cm² do not further affect this density value (190). Quartz, tridymite, and cristobalite attain the same density after heavy neutron irradiation, which means a density decrease of 14.7% for quartz and 0.26% for cristobalite (191). The resulting glass-like material is the same in each case, and shows no x-ray diffraction pattern but has identical density, thermal expansion (192), and elastic properties (193). Other properties are also affected, ie, the heat capacity is lower than that of vitreous silica (194), the thermal conductivity increases by a factor of two (195), and the refractive index, *n*_D, increases to 1.4690 (196). The new phase is called amorphous silica M, after metamict, a word used to designate mineral disordered by radiation in the geological past (197).

A number of mechanisms have been proposed by which this common irradiated state is obtained. The most widely accepted is the thermal spike theory, which considers the heat generated in the wake of a fast particle passing through a solid as being sufficient to cause severe structural disturbances which are then frozen in by rapid cooling. Many property changes can be explained by this theory (146).

The structure of silica M has not, as of this writing (ca 1996), been determined. It is generally agreed that silica M is amorphous, although it exhibits a higher degree of order than does vitreous silica. An electron diffraction pattern resembling that of α-quartz can be obtained from very heavily irradiated vitreous silica (198). No crystallinity was observed in samples given doses below an integrated influx of 8.6 × 10¹⁹ nvt, where *n* is the number of neutrons, *t*

the velocity, and *t* the time of exposure. Crystallinity so induced shows a large number of lattice defects and escapes detection by x-ray diffraction methods. This α-quartz model is further supported by the observation that amorphous silica M changes to polycrystalline quartz when heated at 940°C for 16 h (197). Confirmation of these results is required before this theory can be fully accepted. Radiation compaction of vitreous silica has been reviewed (199), and the effects of electron irradiation have been described (200–202).

Irradiation by both x-rays and γ-rays produces absorption centers in vitreous silica, although structural damage is negligible. These absorption centers show up primarily in the visible and ultraviolet spectral regions. Little change is seen in the infrared region; a Raman line at 606 cm⁻¹ has been ascribed to nonbridging oxygen defects (203). The types and number of absorption centers produced depend on the purity of the vitreous silica and the total radiation dosage received. The rate of coloration of high purity vitreous silica, ie, very low concentration of metallic impurities, such as Type III silica, is very slow, and doses of >10 Gy (10⁷ rad) are required to obtain appreciable absorption. For comparable absorption intensities in less-pure material, ie, silica Types I and II, radiation approximately two orders of magnitude lower is required (204). Extensive research has been conducted in the field of radiation-induced absorption centers in vitreous silica. Excellent reviews, including a thorough compilation of all absorption centers studied, are available (205–207). Only the principal centers and their probable causes are discussed herein.

A typical absorption curve for vitreous silica containing metallic impurities after x-ray irradiation is shown in Figure 12. As shown, the primary absorption centers are at 550, 300, and between 220 and 215 nm. The 550-nm band results from a center consisting of an interstitial alkali cation associated with a network substituent of lower valency than silicon, eg, aluminum (205). Only alkalis contribute to the coloration at 550 nm. Lithium is more effective than sodium, and sodium more effective than potassium. Pure silica doped with aluminum alone shows virtually no coloration after irradiation. The intensity of the band is determined by the component that is present in lower concentration. The presence of hydrogen does not appear to contribute to the 550-nm color-center production (209).

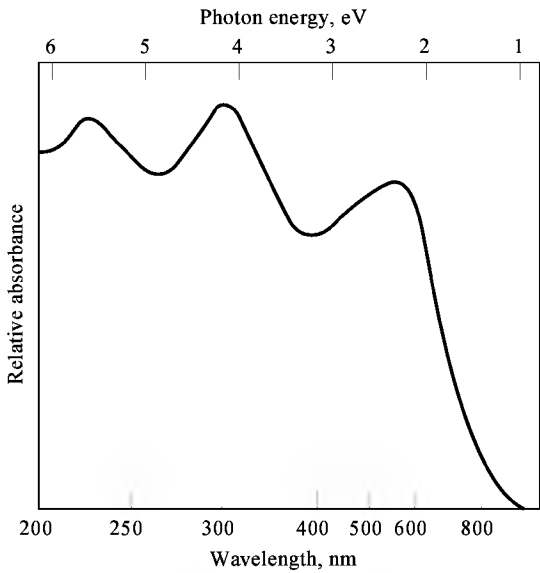


Fig. 12. Absorption spectrum of irradiated impure vitreous silica, Heraeus fused quartz, after a radiation dose of 10⁴ Gy (10⁷ rad) (208). The main impurity is aluminum.

The absorption band at 300 nm may also be associated with alkali ions, possibly the result of a trapped electron stabilized by an alkali ion. The band shifts to longer wavelengths when heavier alkali ions are present, and growth rates for the band show a definite dependence on the type of alkali (205,209).

The 215-nm band may be intrinsic to silica. This band can be produced in Corning Code 7940 glass by long-term x-ray irradiation (210). This band is attributed to an E' center, which may also be observed in irradiated α-quartz. Structurally, the E' center is assumed to be a pyramidal SiO₃ unit having an unpaired electron in the dangling *sp*³ orbital of Si (207).

A typical absorption curve obtained for a metal-free vitreous silica after a large dose of γ-rays is shown in Figure 13. The main band is at 215 nm; three smaller bands occur at 230, 260, and 280 nm. The 230-nm band may result from an electron trapped at a silicon atom having an incomplete oxygen bond (205).

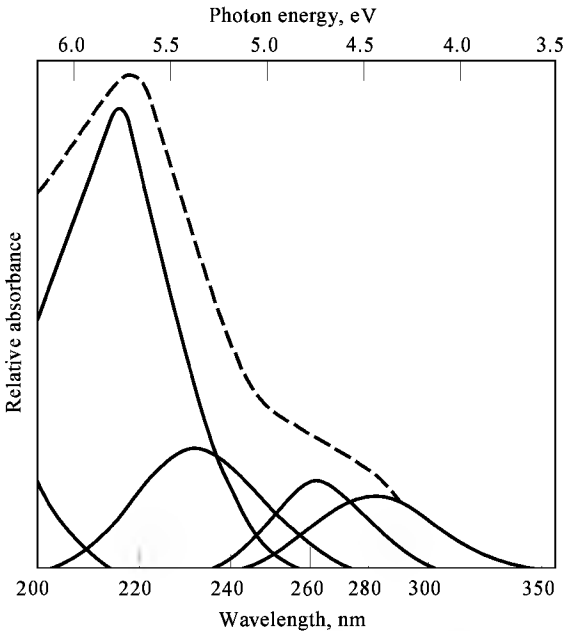


Fig. 13. Total (—) absorption spectrum and resolved (---) bands of irradiated high purity vitreous silica (211). Measurements made on Corning Code 7940 glass after a radiation dose of 26 Gy (2600 rad).

The band at 280 nm has been shown to be a germanium center (207). Nonbonding oxygens having negative charge can act as hole traps in irradiated silica. These oxygen-associated hole centers (OHCs) have been associated with an absorption at 163 nm. Dry OHCs, observed in vitreous silicas having low hydroxyl content, are peroxy-radical defects (212).

Excimer lasers (qv) generate pulsed light having very high instantaneous peak powers. These lasers operate in the deep ultraviolet (Duv) at 308, 248, and 193 nm, and have been found to cause laser-induced red fluorescence (658 nm), induced absorption, and structural compaction of vitreous silica. The induced absorption at 215 nm is the same E' defect center generated using more energetic beams such as γ- or x-ray. The subband gap light generated by excimer lasers excites electrons into the valance band by a two-photon process. Transmission measurements made as a function of intensity show that the two-photon absorption (TPA) coefficient is approximately $(1.9\text{--}3.0) \times 10^{-10}$ cm² W at 193 nm and 4.5×10^{-11} cm² W at 248 nm (213). The compaction and color-center formation also vary quadratically with the laser fluence or intensity and linearly with the number of pulses. The compaction causes a geometrical path decrease in the material along with an increase in the index of refraction (214). The induced absorption and compaction by excimer irradiation has been shown to reach a saturation point at about 0.1% cm (215).

Economic Aspects

The market for fused silica started in 1906 with the sale of silica muffles and pipes. That same year resulted in the incorporation of the Thermal Syndicate Ltd. Since that time, worldwide sale of vitreous silica material and fabricated products has continued to grow. The sales of vitreous silica ingots, tubes, rods, plates, fabricated products, photomask blanks, crucibles, and optics was estimated to be between \$800 million to \$1 billion in 1995. These figures do not, however, take into account the optical waveguide market based on fused silica technology.

The optical application market excluding optical waveguide uses was estimated at \$14.6 million in the United States in 1994 (216). The annual U.S. consumption of fused quartz has been estimated at close to 45,000 metric tons. Prices ranged from \$50 to \$60 kg.

Prices for individual products vary greatly with volume, purity, and optical quality. In 1994 the price for a 1000-mL fused quartz beaker was about \$100, whereas the prices for a 10 × 10 × 5 cm³ plate ranged from \$125 to \$5000, depending on the optical homogeneity, inclusions, and transmission. Commercial-grade tubing, 25–30-mm OD, sold for \$45–\$90 m and ultrapure 10-mm rod for \$375 m (\$2160 kg).

In constant dollars, many items decreased in cost in the 1990s relative to the early 1980s because of improved manufacturing methods, new raw material sources, and higher production volumes. However, many products have been developed since the mid-1980s, eg, optics for photolithography and excimer lasers, whereas others have been practically displaced by newer materials or technologies. The illumination and projection optics for optical lithography are an emerging market for synthetic fused silica. The stepper and scanner technology used for the production of integrated circuits (qv) is shifting to shorter wavelength light sources in order to improve resolution. These deep uv wavelengths required synthetic fused silica optics. The annual growth rate for the projection optics is estimated at 44% yr for stepper scanner (217). In addition, integrated circuit manufacturing is increasing silicon wafer sizes from 200 mm to 300 mm in diameter, thereby creating a significant increase in demand for high purity fused silica tubing for diffusion furnaces, and silicon wafer furniture. The principal U.S. and foreign manufacturers of vitreous silica are listed in Table 10.

Table 10. Manufacturers and Suppliers of Vitreous Silica

Company	Location
United States	
A. A. I. Products Inc.	Woodbridge, N.J.
Cabot Corp.	Boston, Mass.
Corning, Inc.	Corning, N.Y.
Dynasil Corp. of America	Bedin, N.J.
GE Quartz Products	Cleveland, Ohio
OSRAM-Sylvania	Exeter, N.H.
Heraeus Amersil Inc. ^a	Duluth, Georgia
J. P. Stevens & Co., Inc.	Greenville, S.C.
Pyromatics, Inc.	Willoughby, Ohio
Quartz Products Co. ^b	Louisville, Ky.
Quartz Scientific, Inc.	Fairport Harbor, Ohio
Outside United States	
Asahi Glass	Tokyo, Japan
Heraeus Quarzglass GmbH	Hanau, Germany
Hitachi Cable	Tokyo, Japan
Iruvisil Co. Ltd.	St. Petersburg, CIS
Fujikura	Tokyo, Japan
Mitsubishi Cable	Itami, Japan
Shin-Etsu Quartz Products Co., Ltd.	Kokai, Japan
SICO	Jena-Burgau, Germany
Sumitomo Electric Industry	Osaka, Japan
Thermal Quarz Schmelze GmbH ^a	Weisbaden, Germany
Thermal Syndicate, Ltd. ^b	Wallsend, U.K.
Toshiba Ceramics	Tokyo, Japan
TOSOH Nippon Silica Glass	Tokyo, Japan
Quartz & Silice ^b	France, Holland, Italy
Westdeutsche Quarzschmelze GmbH ^c	Munchen, Germany

^a Subsidiary of Heraeus Quarzglass GmbH.
^b Subsidiary of Saint-Gobain Group Co. Quartz Technology Division.
^c Subsidiary of General Electric.

Uses

Chemical Applications. Because of its excellent chemical durability, high purity, thermal shock resistance, and usefulness at high temperature, vitreous silica has a wide range of applications in chemical analysis and preparations. Tubing, rods, crucibles, dishes, boats, and other containers and special apparatus are available in both transparent and nontransparent varieties (119,137,138,218). Because of its inertness, vitreous silica is used as a chromatographic substrate in the form of microparticles, capillary tubing, and open columns for high resolution gas chromatography (see ANALYTICAL METHOD§; CHROMATOGRAPHY) (219).

Thermal Applications. The protection of precious-metal thermocouples in high temperature pyrometry is an important application of vitreous silica. Although satin tubing is usually employed, transparent tubes are superior for protecting couples when used in a reducing atmosphere (220).

Vitreous silica is used for gas-heated or electrically heated devices in various shapes, eg, as a tube or muffle because of its electrical resistivity, impermeability, and low expansion. In its simplest form, an electric-resistance furnace consists of a vitreous silica tube or pipe on which the resistance element is wound (see FURNACES, ELECTRIC). Because of its indifference to temperature gradients, a tubular furnace of vitreous silica may be made to operate at different temperatures at various portions of the tube, either by arrangement of the heating elements or by cooling sections of the tube with water. Vitreous silica pipes may be employed in vacuum-induction and gas-fired furnaces (see VACUUM TECHNOLOGY) (221).

Radiant heaters employing resistance wire encased in vitreous silica tubes have the capability to modify the emitted radiation and furnish a higher proportion of shorter (1–2 μm) wavelengths which constitute the desirable radiations. Immersion heaters for use with acid solutions are of similar construction. An overhead heating unit consisting of a resistance wire sealed inside a vitreous silica container permits acid liquids to be concentrated or evaporated without ebullition or spattering. High efficiency radiators approximately 200-cm long have been developed for power requirements up to 6 kW. These are used as energy sources in heat exchangers (see HEAT EXCHANGE TECHNOLOGY), high duty enamel drying equipment, polymerization processes, and copying devices (221).

Insulation having high thermal endurance has been made from vitreous silica fibers (see ABLATIVE MATERIALS). Such material forms the basis for the 30,000 insulating tiles, 125-mm thick, that protect the aluminum skin of the space shuttle. The tiles add a minimum of weight because the density of the insulation is only 144 kg m³ (9 lb ft³), similar to balsa wood.

Optical Applications. Vitreous silica is ideal for many optical applications because of its excellent ultraviolet transmission, resistance to radiation darkening, optical polishing properties, and physical and chemical stability. It is used for prisms, lenses, cells, windows, and other optical components where ultraviolet transmission is critical. Cuvettes used in scatter and spectrophotometer cells are manufactured from fused silica and fused quartz because of the transmissive properties and high purity (222).

The development of high power laser systems at shorter uv wavelengths requires the transmissive and optical properties of fused silica. These high power lasers are utilized for commercial energy research and military applications. Relatively new commercial excimer lasers are utilized in photoablation, photolithography, material processing, and medical application. The optical beam delivery systems for these excimers can utilize fused silica lenses or optical fibers (223) as well as fluoride-based optics such as CaF₂ and MgF₂.

All higher energy laser systems throughout the world utilize a significant amount of high quality fused silica. Inertial confinement research laser systems, such as the Lawrence Livermore National Laboratory (LLNL) NOVA laser which operates at 366 nm (frequency tripled Nd–YAG laser), require fused silica for the final focal lens and the debris shield (224). Inertia confinement fusion research is also underway using KrF (248-nm) laser systems such as Los Alamos Aurora and Naval Research Laboratory NIKE. These utilize a fused silica (Coming Code 7940). The Aurora amplifier window was manufactured from a 1-m² fused silica blank (225). The LLNL atomic vapor laser isotope separation (AVLIS) system used for uranium isotope separation experiments utilizes fused silica throughout the optical system.

Mechanical Applications. The volume of vitreous silica used for fibers is a very small part of the total consumption. However, some interesting and significant applications have been developed in the laboratory, particularly in the area of measurements.

The sorption balance employs 15 small vitreous silica springs enclosed in glass tubes (226). The fiber is 2 mm in diameter and the coils 1.25 cm in diameter and 5.99 cm long (see WEIGHING AND PROPORTIONING). A platinum bucket, which weighs 202 mg, contains up to 0.5-g charcoal; the balance can be read to 0.2 mg. The sorption of gases is determined over a wide range of temperatures and pressures. This type of balance has been used for density determinations, measurements of heat loss and evaporation, study of chemical reactions between gas and solid phases, and weight determination of wet tissues of plant and animal origin. The advantages of vitreous silica are its resistance to corrosion, ease of cleaning and sterilization in an enclosed tube, and absence of damping from internal friction.

Because of its low and regular thermal expansion, vitreous silica is employed in apparatus used to measure the thermal expansion of solids. A detailed account of the different methods used for this purpose has been published (227). The most common form of dilatometer utilizes a vitreous silica tube closed at the bottom and containing the test sample. A movable rod of vitreous silica, resting on the sample, actuates a dial indicator resting on the top of the rod. The assembly containing the sample is placed in a furnace, bath, or cooling chamber to attain the desired temperature.

Lighting. An important application of clear fused quartz is as envelop material for mercury vapor lamps (228). In addition to resistance to deformation at operating temperatures and pressures, fused quartz offers ultraviolet transmission to permit color correction. Color is corrected by coating the inside of the outer envelope of the mercury vapor lamp with phosphor (see LUMINESCENT MATERIALS). Ultraviolet light from the arc passes through the fused quartz envelope and excites the phosphor, producing a color nearer the red end of the spectrum (229). A more recent improvement is the incorporation of metal halides in the lamp (230,231).

Incandescent tungsten–halogen (F, Br, I) cycle lamps first became available in 1959 for general use. In this lamp, tungsten evaporating from the filament deposits on the envelope wall where it reacts with halogen vapor to form a volatile halide. The tungsten halide diffuses back to the filament where it dissociates to halogen vapor and tungsten. This permits a higher temperature operation, which affords higher efficiency and longer life than the conventional incandescent lamp. The wall temperature must be at least 250°C and, if possible, around 600°C. Fused quartz has been essential for this development because it is one of the few readily available transparent materials that can be used in lamp envelopes. Improvements include the internal deposition of a barrier, eg, aluminum fluoride or aluminosilicate (232,233), and reflective layers, eg, titanium silicate or zirconia (234,235). These protect the silica glass from vapor attack, slow the diffusion of deactivating impurities, and reflect heat onto the filament while transmitting visible light. These lamps are used in the illumination of airfields, sports arenas, buildings, streets, parking lots, and automobiles. They are also used in slide and film projectors and specialized optical instruments (236,237).

Electronic Applications. In electronic systems, such as radar and computers, signal delay is sometimes necessary. A transducer converts electrical signals to ultrasonic elastic waves, which pass through a connecting medium to another transducer where the waves are reconverted to electrical signals. Vitreous silica is an ideal connecting medium because it has excellent physical stability and low ultrasonic transmission losses. It transmits an ultrasonic signal almost 100,000 times more slowly than an electric signal in a wire. The vitreous silica delay line is in the form of a flat polyhedral plate. The facets of the edges of this plate are ground to a predetermined angle with great precision. The transducer is fastened to one of these facets. The plates may be designed for path lengths of 4 cm to nearly 12 m, giving delay times for shear waves of 10–3000 μs (238).

By far the most significant electronic application, and perhaps the highest volume application for vitreous silica, is in semiconductors (qv). This association began with silicon wafer technology, where vitreous silica was the principal high purity material for crucibles for silicon crystal growth. Both opaque sintered and transparent vitreous silica from tubing or boules are used. Crucible sizes have been increasing in response to the demand for ever-larger silicon wafers.

Vitreous silica tubes are used for the subsequent high temperature diffusion, doping, and epitaxial growth treatments performed on wafers destined for large-scale integrated circuits. Once again, the high purity of certain grades of vitreous silica, along with its high thermal endurance, make it the material of choice for diffusion tubes, wafer holders, boats, and associated hardware. Quartz manufacturers continue to push for higher purity, ie, alkali content in tubing approaching <1 ppm. The thermal endurance of the tubes may be prolonged by causing some surface crystallization; the increased rigidity prevents sagging (137). Diffusion-resistant barrier coatings have also been used (239,240). In addition to silicon wafers, many other high temperature, high purity solid-state reactions are performed in vitreous silica vessels, including crystal growth, zone refining, and the preparation of gallium and indium arsenides as well as germanium and silicon optical materials.

Thin films (qv) of vitreous silica have been used extensively in semiconductor technology. These serve as insulating layers between conductor stripes and a semiconductor surface in integrated circuits, and as a surface passivation material in planar diodes, transistors, and injection lasers. They are also used for diffusion masking, as etchant surfaces, and for encapsulation and protection of completed electronic devices. Thin films serve an important function in multilayer conductor insulation technology where a variety of conducting paths are deposited in overlay patterns and insulating layers are required for separation.

Thin vitreous silica films are usually formed by vapor deposition or r-f sputtering (see THIN FILMS, FILM DEPOSITION TECHNIQUES). Vapor deposition is generally effected by the pyrolytic decomposition of tetraethoxysilane or another alkoxysilane. Silica has been most extensively used in r-f sputtering of

dielectric films, which are of very high quality.

Large-scale and very large-scale integration (VLSI) of electronic circuits requires the photoreduction of complex conductor and insulator patterns, which are reproduced on semiconductor devices by various photoresist processes. Vitreous silica is frequently used for the photomask substrate, ie, the transparent substrate for the image mask that contains the basis for conduction or insulator patterns. Because of the extremely high dimensional tolerances of VLSI, vitreous silica is a natural choice for the substrate. Its uv transmission properties allow fast exposure of the photoresists, and its very low thermal expansion prevents pattern distortions owing to temperature gradients encountered during processing. Perfectly flat and inclusion-free material is required (241). Another application for fused silica in the semiconductor field is as optical elements in microlithographic systems. In order to obtain smaller features on computer chips, ie, increasing the memory capacity, photolithographic stepper and scanner systems are moving to shorter wavelengths. Vitreous silica is used to make the excimer-based deep ultraviolet (248 and 193 nm) lithographic lens systems because of its superior uv transmission and radiation resistance characteristics. These lens systems are used to manufacture 64 Mb and 256 Mb dynamic random-access memory computer chips (242).

Space and Astronomy. Vitreous silica is used in several space-based applications because of static fatigue (slow crack growth), thermal stability, and radiation resistance. Every U.S. space vehicle having service personnel, including Mercury, Gemini, Apollo, and space shuttle vehicles, has been equipped with windows made of high optical-quality vitreous silica (Corning Code 7940) in order to have the clarity needed for visual, photographic, and television-based observations. The space shuttle utilizes triple-layer windows that have outer and central panes of vitreous silica with a tempered aluminosilicate inner pane. The outer pane is thinner for thermal endurance, whereas the two inner panes are thicker to supply strength (243).

The ability of vitreous silica to withstand dimensional changes with changing temperature has made it ideal for mirror blanks of telescopes (244–246). Vitreous silica is used in both space-based and ground-based mirror applications because of the thermal stability and the ability to polish to ultrasmooth surface. The mirrors for space-based applications are fabricated by fusion-sealing struts into a honeycomb structure of vitreous silica, which is then fused or bonded with expansion-matching materials to top (face plate) and bottom (backer) plates. Weight reductions of up to 90% , relative to the weight of an equivalent-sized solid glass blank, can be achieved using these sandwich structures.

An ultralow expansion (ULE) vitreous silica (ULE-Corning Code 7971) can be manufactured by doping the silica with a 7.5% titanium dioxide (144,243). The sonic velocity and coefficient of thermal expansion (CTE) are directly related to the titanium concentration. This material was used to manufacture the 2.4 m dia lightweight mirror blank for the Hubble telescope and the 8.3 m mirror blank for the ground-based Japanese National Large Telescope (JNLT). The JNLT piece, the largest mirror blank in the world, was manufactured by hex-sealing titanium-doped vitreous silica pieces into a solid glass mirror weighing >30 t (247).

Space-based solar cells are covered with a very thin layer of vitreous silica to protect against the damaging environment of space such as atomic oxygen, micrometeorites, and radiation effects. Because the silica is transparent to damaging uv radiation, it is normally coated with a uv-reflective thin film (see SOLAR ENERGY).

The Topex oceanographic satellite used a laser-based retroreflector array for positioning. The retroreflectors were manufactured from Corning Code 7958 fused silica, a sol–gel-derived low water vitreous silica material (248).

Optical Fibers. Pure and doped fused silica fibers have replaced copper lines in the telecommunication area. Fused silica fibers are used in laser surgery, optical sensor application, and laser welding (see SENSORS). Optical-fiber-tethered weapons such as fiber-optics-guided (FOG) missiles are another potential application for fused silica (249,250) (see FIBER OPTICS).

Other Uses. Vitreous silica also has applications in the form of powders, fibers, wool, and chips (251). The powder may be used as an inert filler or as a coating material in the investment-casting industry for the lost-wax process. Chips are used as an inert substrate in certain chemical process applications. Vitreous silica wool or fiber is an excellent insulation or packing material. Complex shapes rendered in extruded or sintered vitreous silica are used to produce precisely shaped holes, tunnels, cavities, and passages in investment-cast metal parts (see also REFRACTORIES). The vitreous silica shapes are chemically compatible with most alloys and dimensionally stable at casting temperatures. They give smooth internal surfaces to the cast pieces. The cooled casting is subsequently treated in hot caustic solutions to remove the silica core. Sandblasting may also be used. Internal features such as long holes and cooling passages for turbine rotors, which would be impossible to fabricate by drilling or machining, are obtained routinely (252).

Antireflective (AR) coatings are required on optics to reduce the reflective surface losses. Vitreous silica coatings in the form of porous or multilayer films are used extensively in this application. Antireflective coatings have been developed which employ colloidal fused silica sol–gel particles made from organometallic materials (253).

BIBLIOGRAPHY

"Vitreous Silica" under "Silica and Silicates" in *ECT* 1st ed., Vol. 12, pp. 335–344, by W. Winship, The Thermal Syndicate Ltd.; "Silica, Vitreous" in *ECT* 2nd ed., Vol. 18, pp. 73–105, by W. H. Dumbaugh and P. C. Schultz, Corning Glass Works; in *ECT* 3rd ed., Vol. 20, pp. 782–817, by P. Danielson, Corning Glass Works.

1. F. E. Wagstaff, *J. Am. Cer. Soc.* **52**, 650 (1969).
2. R. B. Sosman, *The Phases of Silica*, Rutgers University Press, New Brunswick, N.J., 1965, pp. 164–165.
3. B. J. Skinner and J. J. Fahey, *J. Geophys. Res.* **68**(19), 5595 (1963).
4. Ref. 2, pp. 148–150.
5. J. S. Laufer, *J. Opt. Soc. Am.* **55**, 458 (1965).
6. G. Hetherington, *J. Brit. Ceram. Soc.* **3**(4), 595 (1966).
7. I. Fanderdik, ed., *Silica Glass and its Application*, Elsevier Science Publishers, New York, 1991, pp. 165–166.
8. G. Hetherington, K. H. Jack, and J. C. Kennedy, *Phys. Chem. Glasses*, **5**(5), 130–136 (1964).
9. D. N. Casey, G. Hetherington, J. A. Winterburn, and B. Yates, *Phys. Chem. Glasses*, **17**(3), 77–81 (1976).
10. L. L. Hench, S. H. Wang, and J. L. Nogues, *Multifunctional Materials*, SPIE Vol. 878, Society of Photooptical Engineers, Bellingham, Wash., 1988, pp. 76–85.
11. D. Torikai and co-workers, *J. Non-Cryst. Solids*, **179**, 328–334 (1994).
12. R. L. Mozzi and B. E. Warren, *J. Appl. Cryst.* **2**, 164–172 (1969).
13. E. Lorch, *J. Phys. C*, **2**, 229 (1969).
14. A. C. Wright, *J. Non-Cryst. Solids*, **179**, 84–115 (1994).
15. D. I. Grimley, A. C. Wright, and R. N. Sinclair, *J. Non-Cryst. Solids*, **119**, 49 (1990).
16. D. L. Price and J. M. Carpenter, *J. Non-Cryst. Solids*, **92**, 153 (1987).
17. J. R. G. de Silva, D. G. Pinatti, C. E. Anderson, and M. L. Rudee, *Phil. Mag.* **31**, 713 (1975).
18. V. E. Sokol'skii, V. A. Shovskii, V. P. Kazimirov, and G. I. Batalin, *Fizika I Khimiya Stekla*, **6**(5), 517–520 (1980).
19. R. J. Bell and P. Dean, *Phil. Mag.* **25**, 1381 (1972).
20. A. E. Geissberger and F. L. Galeener, *Phys. Rev. B*, **28**(6), 3266–3271 (1983).
21. J. C. Mikkelsen and F. L. Galeener, *J. Non-Cryst. Solids*, **37**, 71–84 (1980).
22. E. A. Porai-Koshits, *J. Non-Cryst. Solids*, **73**, 79 (1985).
23. J. C. Phillips, *Phys. Rev. B*, **32**, 5350 (1985); **33**, 4443 (1986).
24. D. L. Griscom, *J. Non-Cryst. Solids*, **73**, 51–77 (1985).
25. F. L. Galeener, in R. A. Weeks and D. L. Kinser, eds., *Effects of Modes of Formation on the Structure of Glasses*, Trans Tech Publications Ltd., Brookfield, Vt., 1987, pp. 305–314.
26. R. Bruckner, *J. Non-Cryst. Solids*, **5**, 123–175 (1970); **5**, 177–216 (1971).

27. D. L. Griscom, *J. Cer. Soc. (Japan)*, **99**(10), 923–942 (1991).
28. G. C. Escher, *Excimer Beam Application* SPIE Vol. 998, Society of Photooptical Engineers, Bellingham, Wash., 1988, pp. 30–37.
29. J. E. Shelby, *J. Non-Cryst. Solids*, **179**, 138–147 (1994).
30. K. Awazu and co-workers, *J. Appl. Phys.* **69**(4), 1849–1852 (1991).
31. T. E. Tsai, D. L. Griscom, and E. J. Friebele, *Phys. Rev. B.* **38**(3), 2140–2146 (1988).
32. D. L. Griscom, M. Stapelbroek, and E. J. Friebele, *J. Chem. Phys.* **78**(4), 1638–1650 (1983).
33. Ref. 2, pp. 96–97.
34. C. R. Gaudin, *Acad. Sci. (Paris)* **8**, 678, 711 (1839).
35. Ref. 2, p. 809.
36. W. A. Shenstone, *Nature* **61**(1588), 540 (1900).
37. W. A. Shenstone and H. G. Lacell, *Nature* **62**(1592), 20 (1900).
38. W. C. Heraeus, *Zeitschr. Electrochem.* **8**, 861–862 (1902); **9**, 847–850 (1903).
39. **U.S. Pat. 2,268,589** (1942), J. A. Heany (to Heany Industrial Ceramics Corp.).
40. **U.S. Pat. 2,272,342** (1942), J. F. Hyde (to Corning Glass Works).
41. **Brit. Pat. 10,670** (1904), J. F. Bottomley, R. S. Hutton, and R. A. S. Paget.
42. **Brit. Pat. 10,437** (1904), J. F. Bottomley and R. A. S. Paget.
43. **U.S. Pat. 2,155,131** (Apr. 18, 1939), W. Hanlein (to Patent-Treuhand-Gesellschaft für Elektrische Glühlampen GmbH).
44. **U.S. Pat. 3,764,286** (Oct. 9, 1973), S. Antczak and co-workers (to General Electric Co.).
45. **U.S. Pat. 2,904,713** (Sept. 15, 1959), W. H. Heraeus and H. Mohn (to Heraeus Quarzschmelze GmbH).
46. **U.S. Pat. 3,128,166** (Apr. 7, 1964), H. Mohn (to Heraeus Quarzschmelze GmbH).
47. **U.S. Pat. 2,270,718** (Jan. 20, 1942), F. Skaupy and G. Weissenberg.
48. **U.S. Pat. 4,072,489** (Feb. 7, 1978), T. A. Loxley and co-workers (to Sherwood Refractories, Inc.).
49. **U.S. Pat. 3,837,825** (Sept. 24, 1974), T. A. Loxley and co-workers (to Sherwood Refractories, Inc.).
50. **U.S. Pat. 3,775,077** (Nov. 27, 1973), C. A. Nicastro and co-workers (to Corning Glass Works).
51. **Fr. Demande 2,399,978** (1978), T. Izawa, T. Kuwabara, and Y. Masuda (to Nippon Tel.).
52. K. Nassau and J. Shiever, *Amer. Cer. Soc. Bull.* **54**, 1004–1009 (1975).
53. A. Audsley and R. K. Bayliss, *J. Appl. Chem.* **19**, 33–38 (1969).
54. A. D. Pinnow and W. F. Dabby, *Mater. Res. Bull.* **10**(12), 1263–1266 (1975).
55. **U.S. Pat. 5,043,002** (Aug. 27, 1991), M. S. Dobbins and R. E. McLay (to Corning Inc.).
56. **Brit. Pat. 2,245,553** (Aug. 8, 1992), I. G. Sayce, Alan Smithson, and P. J. Wells (to Thermal Syndicate Ltd.).
57. G. D. Ulrich and J. W. Riehl, *J. Colloid Inter. Sci.* **87**(1), 257–265 (1982).
58. B. S. Aronson, D. R. Powers and R. G. Sommer, *Proceedings of Topical Meeting on Optical Fibre Communication*, Washington, D.C., 1979.
59. S. Sudo and co-workers, *Electron. Lett.* **14**, 534 (1978).
60. K. J. Beales and C. R. Day, *Phys. Chem. Glasses*, **21**(1), 5–21 (1980).
61. P. C. Schultz, *Appl. Optics*, **18**(21), 3684–3693 (1979).
62. E. M. Dianov and co-workers, *Sov. Lightwave Comm.* **2**, 79–82 (1992).
63. S. Kobayashi and co-workers, *Appl. Optics*, **14**(12), 2817–2818 (1975).
64. I. Matsuyama and co-workers, *Amer. Cer. Soc. Bull.* **63**(11), 1408–1411 (1984).
65. J. Opitz, *Glastech. Ber.* **60**(4), 133–139 (1987).
66. L. L. Hench, *Cer. Int.* **17**, 209–216 (1991).
67. L. C. Klein, *Glass Ind.* **62**(1), 14–17 (Jan. 1981).
68. R. Clasen, *Glastech. Ber.* **60**(4), 125–132 (1987).
69. G. W. Scherer and J. C. Luong, *J. Non-Cryst. Solids*, **63**, 163–172 (1984).
70. **U.S. Pat. 4,059,658** (Nov. 22, 1977), R. D. Shoup and W. J. Wein (to Corning Glass Works).
71. K. Susa and co-workers, *J. Non-Cryst. Solids*, **79**, 165–176 (1986).
72. E. M. Rabinovich and co-workers, *J. Non-Cryst. Solids*, **82**, 42–49 (1986).
73. S. Wang and co-workers, *Sol-Gel Optics II*, SPIE Vol. 1758, Society of Photooptical Engineers, Bellingham, Wash., 1992, pp. 113–124.
74. **U.S. Pat. 4,789,389** (Dec. 6, 1988), P. M. Schermerhorn, M. P. Teter, and R. V. Vandewoestine (to Corning Glass Works).
75. Personal communication, F. Foster, Applied Laser Systems, Inc., Santa Clara, Calif., 1981.
76. W. G. Houskeeper, *J. Am. Inst. Electr. Eng.* **42**, 954 (1923).
77. O. L. Anderson and C. J. Dienes, in V. D. Frechette, ed., *Non-Crystalline Solids*, John Wiley & Sons, Inc., New York, 1960, pp. 449–486.
78. J. Arndt, R. A. B. Devine, and A. G. Revesz, *J. Non-Cryst. Solids*, **131–133**, 1206–1212 (1991).
79. F. M. Ernsberger, in D. R. Uhlmann and N. J. Kreidl, eds., *Elasticity and Strength in Glass*, Vol. 5, Academic Press, Inc., New York, 1980, pp. 9–12.
80. F. P. Mallinder and B. A. Procter, *Phys. Chem. Glasses* **5**, 91 (1964).
81. C. W. Morey, *The Properties of Glass*, Reinhold Publishing Corp., New York, 1954, pp. 319–320.
82. C. L. Babcock, *Silicate Glass Technology Methods*, John Wiley & Sons, Inc., New York, 1977, pp. 56–62.
83. M. R. Vukcevic, *J. Non-Cryst. Solids*, **11**, 25 (1972).
84. R. K. Iler, in E. Matijevic, ed., *Surface and Colloid Science*, Vol. 6, Wiley-Interscience, New York, 1973.
85. W. Stober, *Kolloid-Z.* **147**, 131 (1956).
86. R. Moseback, *J. Geol.* **65**, 347 (1957).
87. *Fused Silica Code 7940 Data Sheet*, Corning Glass Works, Corning, N.Y., 1992.
88. J. J. Miller and D. L. Eppink, *Leaching of Preformed Ceramic Cores*, Sherwood Refractories, Inc., Cleveland, Ohio, 1977.
89. R. Kleinteich, *Quarzglas Quarzgut Instrumenten Technik*, **9**, 334–339; **10**, 365–369; **11**, 409–417; **12**, 449–454 (1961).
90. Y. Oka and M. Tomozawa, *J. Non-Cryst. Solids*, **42**, 535–544 (1980).
91. Ref. 7, p. 201.
92. D. T. Liang and D. W. Readey, *J. Amer. Cer. Soc.* **70**(8), 570–577 (1987).
93. Ref. 2, p. 146.
94. C. A. Elyard and H. Rawson, *Proceedings of the 6th International Congress on Glass*, Washington, D.C., 1962, pp. 270–286.
95. *Vitreosil*, Thermal American Fused Quartz Co., Montville, N.J., 1966.
96. G. H. A. M. VanderSteen, "Introduction and Removal of Hydroxyl Groups in Vitreous Silica," Ph.D. dissertation, Eindhoven, the Netherlands, 1976, pp. 64–70.
97. D. W. Green and R. H. Perry, eds., *Perry's Chemical Engineer's Handbook*, 6th ed., McGraw-Hill Book Co., Inc., New York, 1984, pp. 3–49.
98. A. deRudney, *Vacuum*, **1**, 204 (1951).
99. N. C. Ainslie, C. R. Morelock, and D. Turnbull, *Proceedings of Symposium on Nucleation and Crystallization in Glasses and Melts*, Toronto, Canada, 1961, pp. 97–107.
100. W. D. Kingery, H. K. Bowen, and D. R. Uhlmann, *Introduction to Ceramics*, John Wiley & Sons, Inc., New York, 1976, p. 87.

101. H. Rawson, *Inorganic Glass-Forming Systems*, Academic Press, Inc., New York, 1967, pp. 48–61.
102. P. P. Bihuniak, *Comm. Amer. Cer. Soc.* **66**(10), C188–C189 (Oct. 1983).
103. F. E. Wagstaff, S. D. Brown, and I. B. Cutler, *Phys. Chem. Glasses*, **5**(3), 76 (1964).
104. A. G. Boganov, V. S. Rudenko, and C. L. Bashina, *Izv. Akad. Nauk SSSR Neorg. Mater.* **2**(2), 363 (1966).
105. F. E. Wagstaff and K. J. Richards, *J. Am. Ceram. Soc.* **48**(7), 382 (1965).
106. **U.S. Pat. 3,370,921** (Feb. 27, 1968), F. E. Wagstaff.
107. D. R. Uhlmann, J. F. Hays, and D. Turnbull, *Phys. Chem. Glasses*, **7**(5), 159 (1966).
108. R. C. Yalman and J. F. Corwin, *J. Phys. Chem.* **61**, 1432 (1957).
109. R. H. Doremus, *J. Phys. Chem.* **80**(16), 1773–1775 (1973).
110. H. Wakabayashi and M. Tomozawa, *J. Amer. Cer. Soc.* **72**(10), 1850–1855 (1989).
111. C. Hetherington and K. H. Jack, *Phys. Chem. Glasses*, **5**(5), 147 (1964).
112. R. H. Doremus, *Glass Science*, John Wiley & Sons, Inc., New York, pp. 121–176 (1973).
113. G. H. Frishat, *Ionic Diffusion in Oxide Glasses*, Trans Tech Publications, Bay Village, Ohio, 1975.
114. J. S. Rothman and co-workers, *J. Am. Cer. Soc.* **65**(11), 578–582 (1982).
115. C. H. Frischat, *J. Am. Ceram. Soc.* **51**(9), 528 (1968).
116. G. S. Nakayama and J. F. Shackelford, *J. Non-Cryst. Solids*, **126**, 249–254 (1990).
117. R. H. Doremus, *Glass Science*, John Wiley & Sons, Inc., New York, 1973, pp. 121–145.
118. J. Kirchhof and co-workers, *J. Non-Cryst. Solids*, **181**, 266–273 (1995).
119. *TAFQ*, Thermal American Fused Quartz Co., Montville, N.J., 1980.
120. J. F. Shackelford, J. S. Masaryk, and R. M. Fulrath, *J. Am. Cer. Soc.* **53**(7), 417 (1970).
121. A. F. Wells, *Structural Inorganic Chemistry*, Oxford University Press, U.K., 1967, p. 785.
122. P. W. Bridgman and I. Simon, *J. Appl. Phys.* **24**(4), 405 (1953).
123. R. Roy and H. M. Cohen, *Nature*, **190**, 798 (1961).
124. E. B. Christiansen, S. S. Kistler, and W. B. Gogarty, *J. Am. Ceram. Soc.* **45**(4), 172 (1962).
125. K. Yamana, M. ToKonami, and A. Nakano, *J. Cer. Soc. Japan*, **98**(4), 311–315 (1990).
126. S. Yamagata and co-workers, *J. Cer. Soc. Japan*, **100**(1), 13–16 (1992).
127. W. Jin, R. Kalia, and P. Vashista, *Phys. Rev.* **B50**(1), 118–131 (1994).
128. J. D. Mackenzie, *J. Am. Ceram. Soc.* **46**(10), 461 (1963).
129. R. J. Hemsley, H. K. Hao, and P. M. Bell, *Phys. Rev. Lett.* **57**, 747–750 (1986).
130. C. Meade, R. J. Hemsley, and H. K. Mao, *Phys. Rev. Lett.* **69**, 1387–1390 (1992).
131. W. Primak, L. H. Fuchs, and D. Day, *J. Am. Ceram. Soc.* **38**, 135 (1955).
132. P. Schermerhorn, *Excimer Lasers: Applications, Beam Delivery Systems and Laser Design*, SPIE Vol. 1835, Society of Photooptical Engineers, Bellingham, Wash., 1992.
133. S.-Y. Hsich and co-workers, *J. Non-Cryst. Solids*, **6**, 37 (1971).
134. C. Hetherington, K. H. Jack, and J. C. Kennedy, *Phys. Chem. Glasses*, **5**, 130 (1964).
135. V. Zandian, J. S. Florry, and D. Taylor, *J. Brit. Cer. Soc.* **90**(2), 59–60 (1991).
136. *Dynasil Fused Silica*, Dynasil Corp. of America, Berlin, N.J., 1992.
137. *Fused Quartz Tubing*, General Electric Co., Cleveland, Ohio, 1990.
138. *Optical Fused Quartz and Fused Silica*, Heraeus-Amersil, Inc., Sayreville, N.J., 1981.
139. E. H. Fontana and W. A. Plummer, *Phys. Chem. Glasses*, **7**, 139 (1966).
140. R. Bruckner, *Glastech. Ber.* **37**(9), 413 (1964).
141. Ref. 7, pp. 218–222.
142. D. W. Bowen and R. W. Taylor, *Ceram. Bull.* **57**(9), 818 (1978).
143. R. Bruckner, *Glastech. Ber.* **37**(10), 459–475 (1964).
144. T. A. Hahn and R. K. Kirby, in M. C. Craham and H. E. Hagy, eds., *Thermal Expansion*, American Institute of Physics, New York, 1972, pp. 13–24.
145. J. W. Berthold and S. F. Jacobs, *Appl. Opt.* **15**(10), 2344 (1976).
146. H. E. Hagy, *Appl. Optics*, **12**(7), 1440–1446 (1973).
147. Ref. 2, p. 311.
148. D. N. Casey and co-workers, *Phys. Chem. Glasses*, **17**, 77–82 (1976).
149. J. C. Lasjaunias and co-workers, *J. Phys. Lett.* **41**, L131–L133 (1980).
150. L. C. K. Carwile and H. J. Hoge, *U.S. Army Technical Report*, 67-7-PR, U.S. Army Natick Laboratories, Natick, Mass., 1966.
151. Ref. 100, p. 627.
152. Ref. 7, pp. 228–230.
153. S. Spinner, *J. Am. Chem. Soc.* **45**, 394 (1962).
154. W. B. Hillig, in J. D. Mackenzie, ed., *Modern Aspects of Vitreous State*, Vol. 2, Butterworth & Co., Washington, D.C., 1962, p. 190.
155. I. Naray-Szabo and J. Ladik, *Nature*, **188**, 226 (1960).
156. J. C. Morley, P. A. Andrews, and I. Whitney, *Phys. Chem. Glasses*, **5**, 1 (1964).
157. W. B. Hillig, *Proceedings of Symposium on the Strength of Glasses and the Means to Improve It*, Union Scientifique Continentale du Verre, Charleroi, Belgium, 1962, p. 206.
158. W. B. Hillig, *J. Appl. Phys.* **32**, 741 (1961).
159. M. L. Hammond and S. F. Ravity, *J. Am. Ceram. Soc.* **46**(7), 329 (1963).
160. S. M. Wiederhorn, *J. Am. Ceram. Soc.* **52**, 99 (1969).
161. J. J. Mecholsky and co-workers, *J. Am. Ceram. Soc.* **57**, 440 (1974).
162. Ref. 7, pp. 210–211.
163. S. M. Wiederhorn, in J. B. Wachtman, ed., *Mechanical and Thermal Properties of Ceramics*, National Bureau Studies Special Publication #303, National Bureau of Standards, Washington, D.C., 1969, p. 217.
164. Annual ASTM Standards, Pt. 17, American Society for Testing and Materials, Philadelphia, Pa., 1978, p. 750.
165. J. H. Westbrook, *Phys. Chem. Glasses*, **1**, 32 (1960).
166. J. D. Mackenzie, *J. Am. Ceram. Soc.* **43**, 615 (1960).
167. C. R. Kurkjian, G. W. Kammlott, and M. Munawar Chaudhri, *J. Amer. Cer. Soc.* **78**(3), 737–744 (1995).
168. M. D. Fagan, *Proc. Natl. Electron. Conf.* **7**, 380 (1951).
169. O. L. Anderson and H. E. Bummel, *J. Am. Ceram. Soc.* **38**, 125 (1955).
170. J. T. Krause, *J. Appl. Phys.* **42**(8), 3035 (1971).
171. E. W. J. Mitchell and E. G. S. Paige, *Phil. Mag.* **1**(8) 1085 (1956).
172. V. Jain, A. K. Varshneya, and P. P. Bihuniak, *Glastech. Ber.* **61**(11), 321–325 (1988).
173. D. W. Shin and M. Tomozawa, *J. Non-Cryst. Solids*, **161**, 203–210 (1993).
174. E. M. Guyer, *Proc. I.R.E.* **32**, 743–750 (1944).

175. A. E. Owens and R. W. Douglas, *J. Soc. Glass Tech.* **43**, 159–178 (1959).
176. G. H. Sigel, Jr., *J. Non-Cryst. Solids*, **13**, 372–398 (1973).
177. H. Mohn, *60 Jahre Quartzglas-25 Jahre Hochvakuumtechnik*, W. C. Heraeus GmbH, Hanau, Germany, 1965, p. 114.
178. G. Hetherington, K. H. Jack, and M. W. Ramsay, *Phys. Chem. Glasses*, **6**(1), 6 (1965).
179. V. Garino-Canina, *Verres Refrac.* **6** 313 (1958).
180. L. Dong and co-workers, *J. Opt. Soc. Am.* **B11**(10), 2106–2111 (1994).
181. R. V. Adams and R. W. Douglas, *J. Soc. Glass Tech.* **43**, 147 (1959).
182. C. J. Parker, technical data, Corning Glass Works, Corning, N.Y., 1956.
183. W. A. Clayton, *Space Aeronaut.*, 129 (June 1963).
184. I. H. Malitson, *J. Opt. Soc. Am.* **55**, 1205 (1965).
185. B. Brixner, *J. Opt. Soc. Am.* **57**(5), 674–676 (1967).
186. R. M. Waxler and G. W. Cleek, *J. Res. NBS-A. Phys. Chem.* **75**(4), 279–281 (1971).
187. G. Hetherington and K. H. Jack, *Phys. Chem. Glasses*, **3**(4), 129–133 (1962).
188. R. Brückner, *Glastech. Ber.* **37**, 459 (1964); **38**, 153 (1965).
189. J. W. Fleming Jr. and J. W. Schiever, *J. Am. Cer. Soc.* **62**(9–10), 526 (1979).
190. E. Lell, N. J. Kreidl, and J. R. Hensler, in J. Burke, ed., *Progress in Ceramic Science*, Vol. 4, Pergamon Press, Inc., New York, 1966.
191. M. C. Wittels and F. A. Sherrill, *Phys. Rev.* **93**, 1117 (1954).
192. I. Simon, *J. Am. Ceram. Soc.* **41**, 116 (1958).
193. G. Mayer and M. Lecomte, *J. Phys. Radium*, **21**, 846 (1960).
194. A. E. Clark and R. E. Strakna, *Phys. Chem. Glasses*, **3**, 121 (1962).
195. A. F. Cohen, *J. Appl. Phys.* **29**, 591 (1958).
196. W. Primak, *Phys. Rev.* **110**, 1240 (1958).
197. Ref. 2, pp. 150, 173–175.
198. S. Weissman and K. Nakajima, *J. Appl. Phys.* **34**, 3152 (1963).
199. W. Primak, *Compacted States of Vitreous Silica*, Gordon & Breach Science Publications, New York, 1975, pp. 83–127.
200. C. B. Norris and E. P. EerNisse, *J. Appl. Phys.* **45**(9), 3876 (1974).
201. W. Primak, *J. Appl. Phys.* **49**(4), 2572 (1978).
202. T. A. Dellin and co-workers, *J. Appl. Phys.* **48**(3), 1131 (1977).
203. F. L. Galeener and co-workers, in Ref. 24, pp. 284–288.
204. G. W. Arnold, Jr., *J. Phys. Chem. Solids*, **13**, 306 (1960).
205. Ref. 190, pp. 3–93.
206. P. H. Caskell and D. W. Johnson, *J. Non-Cryst. Solids*, **20**, 153, 171 (1976).
207. D. L. Griscom, in Ref. 24, pp. 232–252.
208. J. M. Stevels, in V. D. Frechette, ed., *Non-Crystalline Solids*, John Wiley & Sons, Inc., New York, 1960, pp. 412–441.
209. E. Lell, *Phys. Chem. Glasses*, **3**, 84 (1962).
210. A. J. Cohen, *J. Chem. Phys.* **23**, 765 (1955).
211. P. W. Levy, *Phys. Chem. Solids*, **13**, 287 (1960).
212. E. J. Friebele and co-workers, *Phys. Rev. Lett.* **42**(20), 1346 (1979); M. Stapelbroek and co-workers, *J. Non-Cryst. Solids*, **32**, 313 (1979).
213. O. Kittelmann and J. Ringling, *Optics Lett.* **19**(24), 2053–2055 (1994).
214. M. Rothschild and J. H. C. Sedlacek, SPIE Vol. 1848, Society of Photooptical Engineers, Bellingham, Wash., 1992, p. 537.
215. W. P. Leung and co-workers, *Appl. Phys. Lett.* **58**, 551 (1991).
216. *Advanced Optical Materials GB-103R*, Business Communication Co., Inc., Norwalk, Conn., 1990.
217. *Sub-0.5 Micron Lithography*, The Information Network, Williamsburg, Va., 1995.
218. Technical data, Sylvania Emissive Products, Portsmouth, N.H., 1985.
219. S. G. Hurt and co-workers, *Chromatogr. News.*, **8**(2), 32 (1980).
220. R. A. Ragatz and O. A. Hougen, *Chem. Met. Engr.* **33**, 415 (1926).
221. Ref. 177, pp. 174–177.
222. Technical data, NSG Precision Cells, Farmingdale, N.Y., 1992.
223. D. C. Ferranti, *Microelectro. Manufactur. Test.* 18–20 (Apr. 1989).
224. G. Sucki, *Energy Technology Review*, Lawrence Livermore National Laboratories, Livermore, Calif., Apr. May, 1986, pp. 28–33.
225. L. A. Rosocha and L. S. Blair, *Proceedings of International Conference on Lasers*, 1987, pp. 164–175.
226. J. W. McBain and A. M. Baker, *J. Am. Chem. Soc.* **48**, 690 (1926).
227. P. Hidnert and W. Sauder, *Natl. Bur. Std. (U.S.) Circ.* 486 (1950).
228. E. W. Beggs, *Illum. Engr.* **42**, 435 (1947).
229. H. D. Frazer and W. S. Till, *Illum. Engr.* **47**, 207 (1952).
230. D. A. Larson and co-workers, *Illum. Engr.* **58**, 434 (1963).
231. E. C. Martt, L. J. Smialek, and A. C. Green, *Illum. Engr.* **59**, 34 (1964).
232. J. R. Fitzpatrick and V. W. Goddard, *J. Illum. Engr. Soc.* **10**(2), 107 (1981).
233. **Ger. Offen.** **2,524,410** (Dec. 11, 1975), E. M. Clausen (to General Electric Co.).
234. **U.S. Pat.** **3,988,628** (June 13, 1974), E. M. Clausen (to General Electric Co.).
235. **U.S. Pat.** **3,879,625** (Oct. 9, 1973), C. I. McVey and O. M. Uy (to General Electric Co.).
236. J. A. Moore and C. M. Jolly, *Gen. Electr. Corp. J.* **29**, 99 (1962).
237. T. M. Lemons and E. R. Meyer, *Illum. Engr.* **59**, 723–728 (1964).
238. E. B. Shand, *Glass Engineering Handbook*, McGraw-Hill Book Co., Inc., New York, 1958, pp. 354–355.
239. K. M. Eisele and R. Ruthardt, *J. Electrochem. Soc.: Solid-State Sci. Technol.* **125**(7), 1188 (1978).
240. *T07 Stabilized Diffusion Tubes*, Heraeus-Amersil, Sayreville, N. J., 1980.
241. *Photomask Substrates for Microlithography*, Corning Glass Works, Corning, New York, 1980.
242. S. G. Olson and C. Sparkes, *SPIE*, "Optical Laser Microlithography III", **1264** (1990).
243. H. A. Miska, *Engineering Materials Handbook*, Vol. 4, *Ceramics and Glasses*, ASM International, Metals Park, Ohio, 1991.
244. C. J. Parker, *Appl. Opt.* **7**, 740 (1968).
245. C. L. Rathmann, C. H. Mann, and M. E. Nordberg, *Appl. Opt.* **7**, 819 (1968).
246. Ref. 177, pp. 143–149.
247. *Ceramics Ind.* **143**(5), 36–40 (Oct. 1994).
248. R. R. Johnson, *Optics/Photon. News*, **6**(3), 16–23 (Mar. 1995).
249. *Corguide Optical Fiber*, Corning Inc., Corning, N.Y., 1990.
250. T. C. Starkey, *Photonics Spectra*, **23**(3), 119–122 (Mar. 1989).
251. Technical data, Cabot Corp., Boston, Mass., 1974.

252. J. J. Miller and D. L. Eppink, *Leaching of Preformed Ceramic Cores*, Sherwood Refractories, Inc., Cleveland, Ohio, 1977.
253. H. G. Floch and J. J. Priotton, *Ceramic Bull.* **69**(7), 1141–1143 (1990).

Daniel R. Sempolinski
Paul M. Schermerhorn
Corning, Inc.

SYNTHETIC QUARTZ CRYSTALS

Silicon dioxide [7631-86-9], SiO₂, exists in both crystalline and glassy forms. In the former, the most common polymorph is α-quartz (low quartz). All commercial applications of crystalline quartz use α-quartz, which is stable only below ca 573°C at atmospheric pressure. Some of the properties of α-quartz are listed in Table 1.

Table 1. Properties of α-Quartz^a

Property	Value	
	<i>Structural</i>	
crystal class, space group	32 ^b	
lattice constant, nm		
<i>a</i>	0.491	
<i>b</i>	0.540 ^c	
	<i>Optical</i>	
indexes of refraction, Na D line		
<i>n</i> ₀	1.5442	
<i>n</i> _E	1.5533 ^d	
optically active, Na D line		
α, ° mm	27.71	
transmission range, nm	150–3000	
	<i>Electrical</i>	
resistivity, Ωcm	10 ¹⁵	
dielectric constant		
ε ₁ ^T	4.58	
ε ₃	4.70	
piezoelectric coupling coefficient, ° o	10	
piezoelectric constant, FC N ^e		
<i>d</i> ₁₁	23.12	
<i>d</i> ₁₄	727	
	<i>Mechanical</i>	
hardness, Mohs'	7	
thermal conductivity, W (m·K)	6.69–12.13	
acoustic <i>Q</i>	(0.1–3) × 10 ⁶	

^a Many properties are directionally dependent, therefore values listed are indicative only.

^b Trigonal trapezohedral class of the rhombohedral subsystem.

^c Variable, depending on purity.

^d Birefringent; *n*_E − *n*_o = 0.0091.

^e To convert FC N to stat C · dyn, divide by 333 × 10⁸.

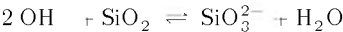
Quartz is mainly used in electronic applications, for which it must be free of electrical and optical twinning, voids, foreign minerals, and liquids. Moreover, it must be large enough for convenient processing. The principal source of electronic-grade natural quartz is Brazil, but manufacturers generally use synthetic quartz for electronic devices. Until the discovery of economic processes for quartz synthesis in the 1950s, natural quartz was used for commercial applications. The size, perfection, and properties of natural quartz vary. In addition, because natural crystals are generally irregular in shape, automated cutting is cumbersome and the yield is lower.

In an attempt to stimulate onshore production of synthetic quartz and piezoelectric devices in the 1970s, Brazil imposed an embargo on exports and ultimately raised the price several-fold for small quartz crystals used as the starting material for quartz growth. However, sources of suitable pure quartz were located in the United States and Canada, including vein and pegmatic deposits (1). Synthetic processes compatible with the natural U.S. quartz starting material from a variety of sources were developed, and U.S. production became relatively independent of imports (1).

As of 1996, synthetic quartz was produced in the United Kingdom, France, CIS, Venezuela, Canada, China, Japan, Brazil, Poland, as well as the United States. The principal nondomestic source is Japan. Some producers in the United States are Eastlake, in Ohio; Motorola, in Carlisle, Pennsylvania; and Thermo Dynamics, in Shawnee Mission, Kansas.

Synthesis

α-Quartz cannot be crystallized from its pure melt because viscous SiO₂ melts almost always form silica glass upon cooling. When crystals are formed, these are high temperature polymorphs of SiO₂, eg, cristobalite [14464-46-1] and tridymite [15468-32-3], that do not easily transform to untwinned α-quartz. α-Quartz is soluble in a variety of molten salts but the melts are viscous and crystallization below the α-transition temperature is not practical. Silicon dioxide is insoluble in most aqueous solvents at ambient conditions, except in HF solutions, and α-quartz is not the stable solid phase in equilibrium with such solutions. No successful practical vapor transport reaction for α-quartz growth has been discovered. However, α-quartz is stable and soluble in water at elevated temperatures and pressures, ie, under hydrothermal conditions. The solubility at a convenient temperature (400°C) and pressure (ca 17 MPa (25,000 psi)) is only a few tenths of a wt % and is not large enough for crystal growth. However, reactions of the following type take place in basic solutions under hydrothermal conditions.



The result is that the weight solubility of quartz is increased to several percent. α-Quartz is the stable solid phase in equilibrium with such solutions and

large crystals can be grown. This is the basis of the processes used for commercial quartz growth (see [HYDROTHERMAL PROCESSING](#)).

The first known successful attempt to grow quartz crystals hydrothermally was in 1905 (2). Small crystals were formed in a thermal gradient from a sodium silicate solution. In Germany during World War II, quartz crystals were grown in an isothermal system using α -quartz seeds and vitreous silica nutrient (3). After that war, research programs aimed at practical quartz production began (4–7). It soon became apparent that processes depending on the supersaturation caused by the presence of silica glass were impractical. Metastable phases have a higher solubility than stable phases, but the supersaturation caused by their presence persists only as long as they are present. In the case of silica glass, its surface devitrifies rather quickly under hydrothermal conditions and growth on α -quartz seeds ceases. All successful quartz growth processes depend on the supersaturation produced by dissolving small particles of quartz nutrient in a hot region of the high pressure system and crystallizing it onto α -quartz seeds in a cooler part of the system. Thus, it is necessary to employ a solvent in which quartz is the stable solid phase having reasonable solubility and in which the dependence of solubility upon temperature produces an appropriate supersaturation, ΔS , and an appropriate temperature differential, ΔT , between the dissolving and growth zones.

All commercial processes (8–10) use either NaOH (4) or Na_2CO_3 (5) as solvent systems. The dissolving mechanism is similar in both solvents because CO_3^{2-} hydrolyzes to OH^- . Sodium salts are required because insoluble sodium iron silicates form on the steel walls of the high pressure vessels when Na^+ and silicates are present (see [SILICON COMPOUNDS](#)). These compounds are relatively insoluble under hydrothermal conditions and form a protective coat on the walls, allowing the use of unlined, relatively inexpensive vessels. The slope of the solubility–temperature curve is greater in CO_3^{2-} than it is in OH^- . Thus loss of control in CO_3^{2-} with poor quality growth and spontaneous nucleation is more likely than it is in OH^- . The practical solution to this problem in CO_3^{2-} is to use a lower ΔT , which produces a ΔS safely below the poor quality region. Thus in CO_3^{2-} processes the growth rates are usually lower. Indeed, in OH^- the temperature control required is generally less than that in CO_3^{2-} .

The hydroxide process developed by Bell Telephone Laboratories (8,9) is used by most producers (11). It is usually operated at higher pressures than the carbonate process and quartz is grown at faster rates. Less precise temperature controls are needed.

Equipment. A typical commercial quartz-growing autoclave is illustrated in Figure 1. The material of construction for use at 17 MPa (25,000 psi) and 400°C can be a low carbon steel, such as 4140, or various types of low alloy steel. The closure, a modified Bridgman closure, is based on the unsupported area principle (12). That is, the pressure in the vessel is transmitted through the plunger to the steel surfaces which initially are nearly line contacts. Thus, the pressure in the seal surface greatly exceeds the pressure in the vessel because most of the area of the plunger is unsupported. Hydrothermal equipment has been further discussed (10).

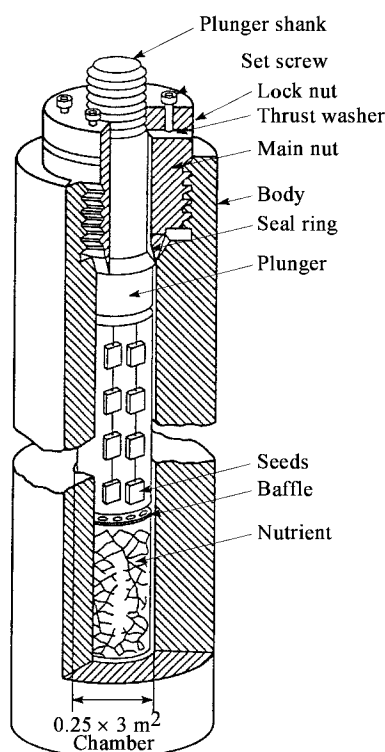


Fig. 1. Hydrothermal quartz autoclave.

After long periods of time at operating temperature, the brittle–ductile transition temperature in autoclave steels increases (13). At temperatures much above 200°C for the solutions and fills used in ordinary hydrothermal processes, pressures and hence stresses in autoclaves can cause failure of metal in the brittle state. Ordinarily, the brittle region is well below these temperatures but careful monitoring of the brittle–ductile transition of the steel is necessary for safe autoclave use over many years.

The baffle is a perforated metal disk that restricts convective circulation within the vessel and thus creates two isothermal regions, the dissolving and growth zones. As a result, all seed crystals experience the same ΔT and ΔS and grow uniformly. No growth takes place on the nutrient. The autoclave is heated by appropriately placed external resistance heaters. The temperature is measured and controlled from externally placed thermocouples; the pressure is measured by strain and Bourdon gauges. In order to increase the yield and hence increase the efficiency and cost effectiveness of the synthesis, the size of the autoclave has been increasing. Originally, autoclaves of 15.24 and 25.4 cm dia were used. More recently, autoclaves up to a meter in diameter have been used to produce over 450 kg of synthetic crystal in a single run.

Modern process facilities are computer controlled. Temperature and ΔT are programmed upward during the start of a growth cycle, pressure and temperature are monitored and controlled, and pressure and temperature overshoot alarms and overrides are provided. Such systems also store data from previous runs for correlations with properties or for identical replication of past conditions.

Typical Run. In a typical run, small particle size α -quartz nutrient is added to a large autoclave, a 5% open area baffle is inserted, and a frame holding many seed plates, the principal surface of which is (0001) or an orientation close to that, is placed within the autoclave. The autoclave is filled to 83% of its free volume with 1.0 M NaOH and then closed. The vessel is heated so that the nutrient zone is at a temperature of about 400°C and the growth zone at a temperature of about 350°C. Under these conditions, the pressure is ca 16.5 MPa (24,000 psi). The vessel is held at these conditions for several weeks and then cooled. Once removed and rinsed with distilled water, the grown crystals are ready for processing. A typical growth rate under such conditions would be ca 1.27 mm/d for growth on the basal (0001) planes. Depending on the particular piezoelectric application, the seed orientation is adjusted to produce a grown crystal that can most efficiently be cut into piezoelectric plates of the orientation dictated by the application. The three seed orientations generally used are basal, $\pm 5^\circ$ (rotated $\pm 5^\circ$ from the basal seed), and seed plates rotated 3° off the minor rhombohedral face, which are sometimes called π - or π -face (14). This last orientation, in spite of the fact that its growth rate is about half of basal growth, gives a grown crystal

particularly useful for AT-cut (~35° off the z -axis) piezoelectric resonators (15) and is often used (14).

Effects of Rate Conditions. It is essential for commercial α -quartz crystals to have usable perfection growth at a high rate and at pressure and temperature conditions that allow economical equipment design. The dependence of rate on the process parameters has been studied (8,14) and may be summarized as follows. Growth rate depends on crystallographic direction; the (0001) is one of the fastest directions. Because ΔS is approximately linear with ΔT , the growth rate is linear with ΔT . Growth rate has an Arrhenius equation dependence on the temperature in the crystallization zone:

$$k = e^{-\frac{\Delta E}{RT}}$$

where k is the rate in a particular crystallographic direction, E the energy of activation, T the absolute temperature, and R the gas constant. Growth rate is increased by an increase in fill, ie, percentage of free volume filled with solvent at room temperature.

Piezoelectrical quartz must be free of twins, bubbles, and particulate inclusions. If the seeds used in quartz synthesis are untwinned and the growth takes place at an appropriate rate and pressure–temperature conditions, synthetic quartz is relatively free of the imperfections. If the rate is too high for the pressure–temperature conditions, the growth becomes limited by the rapidity with which silica can diffuse across a locally depleted zone close to the growing seed. The result is imperfect or crevice-flawed growth. Such growth, which contains many voids and liquid inclusions and is not usable for piezoelectric applications, is caused by conditions analogous to those causing dendritic growth in the preparation of metal crystals. Thus the success of commercial processes depends on careful mapping of the pressure, temperature, and temperature differential conditions in order to find regions where crevice flawing does not occur at commercially useful rates.

Particular difficulty with cracking occurs in z -face-grown quartz. High strain in seeds leads to a strain buildup in the grown quartz that exceeds the elastic limit. This can produce cracks in the as-grown material or lead to cracking during subsequent processing. Severe losses in yield can result (14). Careful inspection of seeds between crossed polarizers in an immersion oil of matching index of refraction reveals strain and allows high quality seeds to be selected so that a high yield of unstrained material is possible (14). Strains result from dislocations generated at particulate inclusions of Na–Fe silicates (16). Growth in the absence of iron in silver cans contained in hydrothermal autoclaves greatly reduces inclusions and strain and produces strain-free seeds for subsequent use under normal conditions (16). The pressure in the can is balanced by filling the space between the can and the autoclave with H_2O to a height slightly above that of the NaOH solution used in the can. The higher fill is required to ensure that the can is under compression, a situation less prone to rupture. Careful choice of seeds and conditions using a silver-lined vessel has even resulted in the growth of low dislocation quartz in the laboratory (16).

An additional requirement for piezoelectrical quartz is a high acoustic quality factor, Q (13,17–19). The acoustic Q of a piezoelectric material is equal to the Q of the resistance–capacitance–inductance circuit, which is electrically equivalent to the piezoelectric resonator circuit. The higher the Q of a piezoelectric material, the more efficiently it converts mechanical to electrical energy. In low Q materials, much energy is lost by thermal processes in this conversion. The acoustic Q of ordinary materials may be thought of as being higher the longer they ring when struck mechanically. Lead, for instance, is not used to make bells because it has a low Q .

The acoustic Q s of quartz prepared under a variety of conditions are given in Table 2. The acoustic Q of natural quartz is $(1\text{--}2) \times 10^6$.

Table 2. Acoustic Q of Synthetic Quartz

Synthetic quartz growth solution	Q
1 M NaOH	1×10^5
1 M NaOH + 0.2 M LiOH	$(2\text{--}3) \times 10^5$
1 M NaOH + 0.2 M LiNO ₂	$(0.5\text{--}1) \times 10^6$
1 M NaOH + 0.2 M LiNO ₂ ^a	2×10^6

^a Crystallization in silver-lined tube.

Infrared absorption studies have shown that Q correlates with an absorption at 3 μm associated with an OH-stretching frequency (20). Indeed, infrared absorption provides a useful tool for Q evaluation in rapid production quality control. Infrared and other studies show that Q degradation is caused by proton inclusion in the grown quartz.

The distribution constant for OH or a proton depends on the presence of other ions. The proton enters the quartz lattice interstitially and is compensated by +3 or +2 ions such as Fe^{2+} , Fe^{3+} , and Al^{3+} substituted at Si^{4+} lattice sites. Iron originates in the autoclave and Al^{3+} is an impurity in the natural quartz starting material. Other +1 ions such as Na^+ or Li^+ can also compensate interstitially for non +4 ions at the Si^{4+} site. Thus, Li^+ doping can be used to improve Q as well as quartz growth in a silver liner, or it can be used to exclude iron (21). Furthermore, the distribution constant for impurities depends on the quartz growth rate in a predictable manner, and slower growth excludes protons (22). Systematic studies of distribution constants have been used to identify the commercial conditions used for high Q and high optical transmission quartz growth (22–24). Thus, high Q quartz can be grown at such rates that large crystals (>5 cm grown material) can be produced in 65–90 d.

Health and Safety Factors

The principal consideration in quartz synthesis is the safe management of the high pressures required for growth (see HIGH PRESSURE TECHNOLOGY). The autoclave must be designed on the basis of proper stress analysis and materials selection. Autoclaves are shielded either by armor plate or by placement in cement-lined pits in the ground. Rupture disks that release pressures of ca 35 MPa (5000 psi) above the operating conditions as well as pressure and temperature overrides and alarms are standard practice. Autoclave embrittlement over long service life must be monitored and those in danger of operating at pressures much above ambient in the brittle range should be removed from service. The hydroxide solutions used in growth should be handled with standard safety measures, including eye shields for operators and safety showers. Activities in high pressure areas should be limited to essential maintenance. The disposal of spent solutions presents no special pollution problems because, at the conclusion of growth, runs are not strongly basic and can be easily neutralized.

Uses

The principal uses of α -quartz are mostly related to the piezoelectric effect (see FERROELECTRICS). Crystals are used in electrical filters and oscillators for frequency and timing and frequency-division multiplexing and demultiplexing. They have also been used as electromechanical transducers in ultrasonic generators and electronic weighing microbalances. Crystal quartz is generally the material used wherever precise timing control is required. Thus, it is widely used in computers, portable as well as stationary telephones, citizen band (CB) radios, and watches. The use of quartz surface-acoustic wave devices in cellular phones is an expanding market. The piezoelectric properties and design of filters and resonators have been discussed (15,25). It is estimated that the production of crystal quartz is about 450 metric tons annually (26), but production fluctuates rapidly owing to demand.

The advent of satellites for timing and communication requires the use of quartz that is not only radiation resistant but of greater mechanical strength and accuracy than material used for other purposes. Whereas the amount of quartz required for these, as well as some military uses, is small, exceedingly careful preparation of devices is essential. The global positioning satellites operate in the range of the radioactive Van Allen belt. Because aluminum substituted for quartz has been shown to be detrimental to radiation resistance (20), the concentration of aluminum must be reduced to a few parts per billion in the crystals. This is done by using slow growth and special starting materials low in aluminum, such as previously grown crystals or

crystals grown from high purity fused quartz. The resonator is not fabricated using the AT, but from a doubly rotated stress-compensated (SC) crystal (27), which has been shown to be less temperature sensitive. Use of computerized fabrication equipment has made device preparation possible. Because these devices are subject to large stresses during liftoff into space, they must be free of etch channels which can reduce the mechanical strength of the device. In addition, during the chemical part of the device fabrication, hydrofluoric acid (HF) is used. The HF tends to etch into dislocations and cause etch channels, which not only weaken the device but can affect the frequency stability.

Etch channels can be reduced or essentially eliminated in a number of ways. Use of different orientation of the seed, the so-called *x*-seed, can produce a crystal having almost no dislocations (28). The grown crystal, however, is not the usual size, making resonator fabrication difficult and expensive. Another, more accepted technique involves applying solid-state electrolysis (sweeping). In this procedure (29), blocks are cut from the crystal above and below the seed, leaving the seed to be reused. A voltage of 1000–2000 V cm is then passed across the block in the *c* direction, which causes lithium and hydrogen (or hydroxide) to migrate out of the crystal. This technique is exceedingly effective in reducing etch channels (30).

Quartz also has modest but important uses in optical applications, primarily as prisms. Its dispersion makes it useful in monochromators for spectrophotometers in the region of 0.16–3.5 m. Specially prepared optical-quality synthetic quartz is required because ordinary synthetic quartz is usually not of good enough quality for such uses, mainly owing to scattering and absorption at 2.6 μm associated with hydroxide in the lattice.

BIBLIOGRAPHY

"Synthetic Quartz Crystals" under "Silica" in *ECT* 1st ed., Vol. 12, pp. 331–335, by G. T. Kohman, Bell Telephone Laboratories; in *ECT* 2nd ed., Vol. 18, pp. 105–111, by R. A. Laudise and A. A. Ballman, Bell Telephone Laboratories, Inc.; in *ECT* 3rd ed., Vol. 20, pp. 818–825, by R. A. Laudise and E. D. Kolb, Bell Laboratories.

1. E. D. Kolb and co-workers, *J. Cryst. Growth*, **36**, 93 (1976).
2. G. R. Spezia, *Acad. Sci. Torino*, **44**, 95 (1908).
3. R. Nacken, *Office Technical Service Report*, PB-6498, U.S. Department of Commerce, Washington, D.C., 1945.
4. A. C. Walker and G. T. Kohman, *Trans. Am. Inst. Electr. Eng.*, **67**, 565 (1948).
5. D. R. Hale, *Science*, **108**, 393 (1948).
6. C. S. Brown and co-workers, *Nature*, **167**, 940 (1951).
7. R. A. Laudise, *J. Am. Chem. Soc.*, **81**, 562 (1959).
8. R. A. Laudise, in F. A. Cotton, ed., *Progress in Inorganic Chemistry*, Vol. III, John Wiley & Sons, Inc., New York, 1962, pp. 1–47.
9. A. A. Ballman and R. A. Laudise, in J. J. Gilman, ed., *The Art and Science of Growing Crystals*, John Wiley & Sons, Inc., New York, 1963, pp. 231–251.
10. R. A. Laudise and J. W. Nielsen, in F. Seitz and D. Turnbull, eds., *Solid-State Physics*, Academic Press, Inc., New York, 1961, pp. 149–222.
11. R. A. Laudise and R. A. Sullivan, *Chem. Eng. Prog.*, **55**, 55 (1959).
12. P. W. Bridgman, *Proc. Am. Acad. Arts Sci.*, **49**, 625 (1914).
13. P. L. Key and co-workers, *J. Cryst. Growth*, **21**, 164 (1974).
14. R. L. Barns and co-workers, *J. Cryst. Growth*, **34**, 189 (1976).
15. R. A. Heising, *Quartz Crystals for Electrical Circuits*, D. Van Nostrand, Co., Inc., New York, 1946.
16. R. L. Barns and co-workers, *J. Cryst. Growth*, **43**, 676 (1978).
17. J. C. King, A. A. Ballman, and R. A. Laudise, *J. Phys. Chem. Solids*, **23**, 1019 (1962).
18. A. A. Ballman, R. A. Laudise, and D. W. Rudd, *Appl. Phys. Lett.*, **8**, 53 (1966).
19. A. A. Ballman, *Am. Mineral.*, **46**, 439 (1961).
20. D. M. Dodd and D. B. Fraser, *J. Phys. Chem. Solids*, **26**, 673 (1965).
21. R. A. Laudise, A. A. Ballman, and J. C. King, *J. Phys. Chem. Solids*, **26**, 1305 (1965).
22. N. Lias and co-workers, *J. Cryst. Growth*, **18**, 1 (1973).
23. A. A. Ballman and co-workers, *J. Appl. Opt.*, **7**, 1387 (1968).
24. E. D. Kolb and co-workers, *Mater. Res. Bull.*, **7**, 397 (1972).
25. W. P. Mason, *Piezoelectric Crystals and Their Application to Ultrasonics*, D. Van Nostrand Co., Inc., New York, 1950.
26. A. F. Armington, *Progress in Crystal Growth and Characterization*. Vol. 21, Pergamon Press, New York, 1991, p. 97.
27. H. Merigoux and J. Darces, *Proceedings of the 40th Annual Frequency Control Symposium*, #86CH2330-9, IEEE, Washington, D.C., 1986, p. 140.
28. A. F. Armington and J. J. Larkin, *J. Crystal Growth*, **71**, 799 (1985).
29. B. Fraser, *Physical Acoustics*, Vol. 5, Academic Press, Inc., New York, 1968, p. 54.
30. A. F. Armington and J. F. Balascio, in Ref. 27, p. 230.

Alton F. Armington
Consultant

SILICA, BRICK.

See REFRACTORIES.

SILICATES.

See SILICON COMPOUNDS.

SILICIDES.

See SILICON AND SILICON ALLOYS.

SILICON AND SILICON ALLOYS

Pure silicon,
Chemical and metallurgical,

PURE SILICON

Silicon [7440-21-3], Si, from the Latin *silex*, *silicis* for flint, is the fourteenth element of the Periodic Table, has atomic wt 28.083, and a room temperature density of 2.3 gm cm³. Silicon is brittle, has a gray, metallic luster, and melts at 1412°C. In 1787 Lavoisier suggested that silica (qv), of which flint is one form, was the oxide of an unknown element. Gay-Lussac and Thenard apparently produced elemental silicon in 1811 by reducing silicon tetrafluoride with potassium but did not recognize it as an element. In 1817 Berzelius reported evidence of silicon occurring as a precipitate in cast iron. Elemental silicon does not occur in nature. As a constituent of various minerals, eg, silica and silicates such as the feldspars and kaolins, however, silicon comprises about 28% of the earth's crust. There are three stable isotopes that occur naturally and several that can be prepared artificially and are radioactive (Table 1) (1).

Table 1. Isotopes of Silicon

Isotope ^a	CAS Registry Number	Natural abundance, %	Half-life, s
²⁴ Si	[15759-98-5]		0.10
²⁵ Si	[15759-89-4]		0.22
²⁶ Si	[14932-60-6]		2.23
²⁷ Si	[14276-59-6]		4.14
²⁸ Si	[14276-58-5]	92.23	
²⁹ Si	[14304-87-1]	4.67	
³⁰ Si	[13981-69-6]	3.10	
³¹ Si	[14276-49-4]		2.62 ^b
³² Si	[15092-72-5]		160 ^c
³³ Si	[34809-98-8]		6.1
³⁴ Si	[29675-19-2]		2.8
³⁵ Si	[34809-99-9]		0.9
³⁶ Si	[34810-00-9]		0.5

^a Also known are ²²Si, ³⁷Si, ³⁸Si, and ³⁹Si.
^b Value shown is in hours.
^c Value shown is in years.

Silicon technology has proceeded along two rather diverse paths. The first encompasses metallurgical-grade silicon, a low purity, low cost material used in large quantities in metallurgical applications and as a feedstock for the preparation of the silicon compounds (qv) used in the silicone industry (see SILICON AND SILICON ALLOYS, CHEMICAL AND METALLURGICAL; SILICON COMPOUNDS, SILICONES). The other path encompasses high purity silicon. The latter is also referred to as semiconductor-, electronic-, or transistor-grade silicon, and most impurity levels are in the parts per billion (ppb) range. High purity silicon is much more expensive than the metallurgical grade only because of the additional steps required to remove unwanted impurities. This quality silicon is used primarily in the manufacture of semiconductor devices such as integrated circuits (qv). High purity silicon is used in much smaller quantities than the metallurgical grade and is generally converted into single-crystal form before use. The demand for high purity silicon has risen through the years as the volume of silicon-based semiconductor devices has increased. Estimated usage from 1955 to 1995 was as follows:

year	1955	1965	1975	1980	1985	1990	1995
quantity, t	5–10	50	1,500	2,800	5,000	7,500	13,000

The increase has, however, not been in direct proportion to the increase in the number of semiconductor devices produced, because manufacturing yields have increased dramatically since the silicon transistor became commercially available in 1954 (see ELECTRONIC MATERIALS; SEMICONDUCTORS, SILICON BASED). The chemical properties of silicon are not particularly sensitive to small amounts of impurities and have mostly been determined using low purity material. However, many of the mechanical, electrical, and optical properties are substantially altered by the level of impurities. These have been reexamined in detail since high purity silicon first became available in the late 1940s.

Crystal Structure

At atmospheric pressure, silicon has a diamond cubic structure, ie, two interpenetrating face-centered cubes displaced 1/4, 1/4, 1/4 from each other. When subjected to ca 11 GPa (~110,000 atm) hydrostatic pressure, the diamond structure is converted to a body-centered tetragonal lattice. If a shear component of force is present, the transition pressure is reduced somewhat. Between about 15 and 40 GPa, a simple hexagonal structure is stable; above that, a close-packed hexagonal structure is stable. In addition, a metastable, body-centered cubic form is observed as the pressure is reduced from 10 to 0 GPa (2). Silicon produced at temperatures below about 500°C by vapor deposition is generally amorphous. However, when it is heated to a somewhat higher temperature, crystallization occurs. In the early literature a hexagonal form stable at atmospheric pressure and thought to be analogous to the graphitic form of carbon was reported, but its existence has never been verified. Additional crystallographic data is given in Table 2.

Table 2. Structure of Silicon

Parameter	Value, at 101.3 kPa (1 atm)	
crystal structure	diamond cubic	
lattice spacing ^a , pm	357.0 ^b	
atoms per unit cell	8	
space group	Fd3m	
ionic radius ^c , Si ⁴⁺ , pm	42	

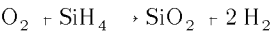
Si–Si single bond length, pm	234
^a Impurity atoms having an ionic radius greater than that of silicon cause lattice expansion.	
^b Value depends on crystal purity. For example, the presence of 0.1 atom % boron causes a lattice constriction of about 0.03%.	
^c Value is based on the Ahrens' calculation.	

Under most circumstances the equilibrium shape of silicon crystals is octahedral, ie, the slowest-growing faces are $\langle 111 \rangle$. However, external conditions can radically alter that shape. For example, when growth is from the vapor, concentration gradients in the gas stream may affect the shape, and when growth is from the melt, the shape is primarily determined by thermal gradients in the melt.

Chemical Properties

Silicon (3), which resembles metals in its chemical behavior, generally has a valence of +4. In a few compounds it exhibits a +2 valence, and in silicides it exists as a negative ion and largely violates the normal valency rules. Silicon, carbon, germanium, tin, and lead comprise the Group 14 (IVA) elements. Silicon and carbon form the carbide, SiC (see CARBIDES). Silicon and germanium are isomorphous and thus mutually soluble in all proportions. Neither tin nor lead reacts with silicon. Molten silicon is immiscible in both molten tin and molten lead.

Silicon forms many compounds that are analogous to those of carbon and thus sometimes appears to give rise to an inorganic equivalent of carbon-based organic chemistry. However, the known silicon compounds (qv) are far fewer and much less complex than those of carbon. One example of these silicon-based compounds is the silane series, eg, SiH₄, Si₂H₆, Si₃H₈, and Si₄H₁₀, which corresponds to the paraffin hydrocarbon series. The silanes react explosively with air and are very poisonous (see SILICON COMPOUNDS, SILANES). Monosilane, SiH₄, however, can be safely decomposed at a few hundred degrees centigrade to give silicon and, provided the concentrations are carefully controlled, can be used to deposit SiO₂ layers on hot surfaces through the following reaction:



Siloxanes are oxygen-containing silanes that appear at first glance to be similar to oxygen-containing hydrocarbon equivalents, but are really much different. Compounds in which silicon directly attaches to an organic radical also form, giving rise to a large number of organosilicon compounds.

Silicon reacts at elevated temperatures with the halogens, forming SiCl₄, SiI₄, SiBr₄, and SiF₄. There is also a series of halogen-substituted silanes such as trichlorosilane, SiH₃Cl, and dichlorosilane, SiH₂Cl₂. Both SiCl₄ and SiH₃Cl are relatively easy to make, purify, and reduce to silicon. These are the silicon compounds most often used as feedstocks in the manufacture of high purity silicon.

Oxygen forms strong bonds with silicon. There are two oxides, quartz or silica, SiO₂, and silicon monoxide, SiO, each of which can exist at room temperature as crystalline or vitreous material; numerous silicates; and almost endless silicone variations. In silica the bond energy is 25.8 kJ (108 kcal) and the bond length 0.163 nm. The nature of the silicon–oxygen bond is of great importance to the semiconductor industry because silicon–oxygen bonding at the oxide–silicon interface of SiO₂-protected silicon devices has a significant effect on silicon semiconductor device performance. Silicon oxidizes to amorphous SiO₂ in an air or oxygen atmosphere, even at room temperature. However, this reaction depends on the diffusion of silicon from the Si–SiO₂ interface to the oxide surface, and at room temperature the oxide thickness is self-limited to 2–3 nm. At higher temperatures the rate of oxide growth (thermal oxide) increases substantially, as shown in Figure 1.

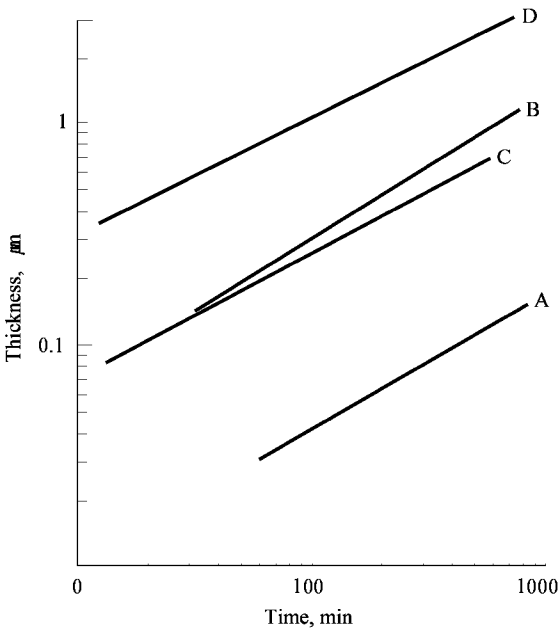


Fig. 1. Thermal oxidation rates for silicon single-crystal $\langle 111 \rangle$ surfaces: A at 900°C in dry O₂; B at 900°C in steam; C at 1200°C in dry O₂; and D at 1200°C in steam.

For the manufacture of silicon semiconductor devices, oxide thicknesses of from <10 to >1000 nm are required on slices of single-crystal silicon. These oxide layers are formed at elevated temperatures, generally at about 1000°C, in an atmosphere of either oxygen or steam. Usually the oxidation is at atmospheric pressure, but sometimes, to speed the oxidation rate, pressures of several atmospheres are used. Oxidation consumes a silicon thickness equal to about 0.4 the thickness of the oxide produced (grown). The thickness of the oxide, X (4) is approximately given by equation 1:

$$X^2 + AX = B(t + \tau)$$
(1)

where A and B are rate constants, τ is a constant involving the initial oxide thickness, and t is the oxidation time. Initially X varies linearly with time, but as oxidation continues, the rate slows and X varies as $t^{1/2}$. The values for A and B depend on temperature, pressure, the oxidant, the crystal orientation of the silicon surface being oxidized, and the amount and kind of impurity in the silicon.

Oxygen also dissolves in the silicon crystal lattice, forming SiO_x, which may radically affect the electrical properties of the silicon. Oxygen is usually unintentionally introduced during the crystal-growing operation in concentrations up to the solubility limit (ca 2.5 × 10¹⁸ atoms cm³). When

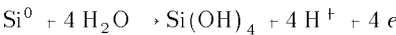
oxygen-containing silicon is annealed in the 450°C range, the OSi complexes become donors and contribute to electrical conductivity. At higher temperatures, larger, electrically inactive clusters form and sometimes cause crystal defects. Above ca 1200°C, the clusters begin to dissolve and the oxygen is redispersed (5). At room temperature and below, silicon appears relatively inert, partly because of the rapid formation in air of a thin layer of protective oxide, SiO₂.

Silicon is virtually insoluble in any single acid but can be dissolved in a two-component, two-stage operation in which one acid component oxidizes the surface and the other etches away the oxide (6). For this purpose, aqueous solutions of HNO₃ and HF are generally used. If a hydrated oxide is formed, eg, by the reaction of silicon and potassium hydroxide or hydrazine, it can be removed with water and a suitable complexing agent (5). Ethylenediamine and isopropyl alcohol are often used with KOH and pyrocatechol with hydrazine. In most proportions the HNO₃–HF–H₂O system etches isotropically and is used for chemically polishing silicon. The others are anisotropic and have very low etch rates on 〈111〉 faces. The anisotropy can be very pronounced. If there are no ledges to promote etching, and if the constituent proportions are properly adjusted, ratios of up to 400 between the 〈111〉 and other faces can be obtained. Anisotropic etching is used in the manufacture of some silicon-integrated circuits and for the micromachining of silicon (7). Etching, again usually using HNO₃–HF–H₂O in various proportions, can be used to delineate crystal defects.

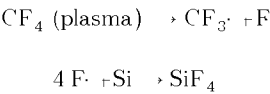
Electrolytic (anodic) etching of silicon occurs in HF solutions (5). Using concentrated HF, the silicon is dissolved in the divalent state:



In dilute HF solutions and higher anodic voltages, the silicon is dissolved in the tetravalent state:



Free radicals such as F· can be used to etch silicon at ambient temperature. The radicals are formed by dissociating compounds such as CF₄ in a plasma. Thus,



Plasma etching is widely used in semiconductor device manufacturing to etch patterns in thin layers of polycrystalline silicon often used for metal oxide semiconductor (MOS) device gates and interconnects (see PLASMA TECHNOLOGY).

Analysis

There are a variety of wet chemistry procedures for analyzing silicon, depending on the silicon concentration and the composition of host material (8). Generally, any silicon is first converted to silica or silicic acid. If the material is organic, this can be done by ashing; otherwise, the use of a fluxing agent such as Na₂CO₃ or Na₂Br₄O₇, followed by an acid treatment to precipitate gelatinous silicic acid, is required. Quantitative analysis by, for example, colorimetry can be done at this point, or the amount of SiO₂ can be determined by the amount of precipitate that dissolves in HF.

Silicon has strong optical emission lines at 251.6113 and 288.1579 nm that can easily be detected by emission spectrography and that give sensitivities in the 1–100-ppm range. For nondestructive analysis, either x-ray diffraction or x-ray fluorescence may be used (see SPECTROSCOPY; X-RAY TECHNOLOGY).

When considering pure silicon, there is much more emphasis on detecting trace impurities in the silicon than on the detection of silicon itself. Whereas optical spectroscopy, secondary ion mass spectroscopy (sims), x-ray fluorescence, neutron-activation analysis, and Auger spectroscopy have all been used, indirect electrical measurements generally provide the greater sensitivity required to detect impurities in the parts per billion (ppb) range. For example, electrical resistivity measurements allow the detection of <1 ppb of an electrically active impurity, although the impurity itself cannot be identified. There are also various measurements that can be made on diodes, eg, deep-level transient spectroscopy (dlts), that allow the presence of impurities in the ppb range to be inferred.

The distribution of impurities over a flat silicon surface can be measured by autoradiography or by scanning the surface using any of the methods appropriate for trace impurity detection (see TRACE AND RESIDUE ANALYSIS). Depth measurements can be made by combining any of the above measurements with the repeated removal of thin layers of silicon, either by wet etching, plasma etching, or sputtering. Care must be taken, however, to ensure that the material removal method does not contaminate the silicon surface.

High Purity Silicon Preparation

Early processes for providing quantities of high purity silicon started with the metallurgical grade of silicon produced by the reduction of silica and coke in an electric arc furnace and the attempts to purify it metallurgically (9). A specific process described as early as 1919 and used during World War II to produce silicon for microwave diodes made use of the fact that most impurities in a melt are segregated at the grain boundaries when the melt freezes. If the frozen polycrystalline agglomerate is crushed, most of the fracturing occurs at the grain boundaries. If the crushed mass is then treated with appropriate acids, many of the impurities that were segregated at the grain boundaries could be leached out. Another related method that was developed in the early 1940s makes use of the fact that the last portion of an ingot to freeze generally has a higher concentration of impurities than the first part. If freezing is constrained to move from one end of the ingot to the other and the ingot is of uniform cross section, the impurity concentration, *C* (10), as a function of *g*, the fraction of ingot solidified, is given by equation 2:

$$C(g) = kC_0(1 - g)^{k - 1}$$

(2)

where *C*₀ is the initial impurity in the melt and *k* is the distribution coefficient of the impurity in question. When *k* is <1, as is generally the case, impurities are concentrated in the last end to freeze, so net purification occurs if that end is cut off and discarded. Remelting the remainder and repeating the process improves purity, but this process is quite unwieldy. Float zone refining, a variant of zone refining (qv) applicable to silicon because it keeps the molten silicon zone from contacting possible contaminants, is shown in Figure 2. This procedure provides a more convenient path to purification by allowing a molten zone of length δ to be moved the length of the ingot, *x*. In that case, instead of equation 2, equation 3 must be used (10).

$$C(x) = C_0 \left(1 - (1 - k) e^{-kx/\delta} \right)$$

(3)

The length of the zone and the diameter of the rod are chosen in such a way that surface tension and interactions between circulating electric currents in the molten zone and the radio-frequency (r-f) field from the surrounding induction coil keep the molten zone in place. As of this writing (ca 1996), the maximum silicon rod diameter that can be purified in this manner is ca 125 mm. Initially, additional purification can be obtained by making more sweeps of the zone. Eventually, however, more sweeps do not remove any additional impurities. The limiting profile is given by equation 4:

$$C(x) = Ae^{Bx}$$

(4)

where *A* and *B* are constants that depend on the value of *k* and the ingot length (10). Values of *k*, the equilibrium distribution coefficient, for various

impurities in silicon are given in Table 3. The k of equations 2 and 3 differs from k_e if freezing is very rapid, but when using float zone refining for purification, the freezing rate is kept low, then $k \approx k_e$.

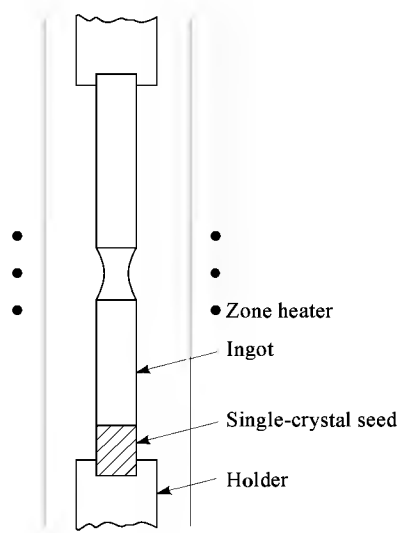


Fig. 2. Float zone purification of silicon.

Table 3. Equilibrium Segregation Coefficients for Impurities in Silicon

Impurity ^a	k_e	Impurity ^a	k_e
Zn	$\sim 1 \times 10^{-5}$	Sb	2.3×10^{-2}
Cd	$\sim 1 \times 10^{-6}$	Bi	7×10^{-4}
B	8.0×10^{-1}	S	$\sim 1 \times 10^{-5}$
Al	2.0×10^{-3}	Cu	4×10^{-4}
Ga	8.0×10^{-3}	Au	2.5×10^{-5}
In	4×10^{-4}	Fe	8×10^{-6}
P	3.5×10^{-1}	O ^b	1.40
As	3×10^{-1}	C	7×10^{-2}

^a Ref. 11.
^b Values <1 have also been reported. Values of k_0 for oxygen in silicon are under extensive discussion.

To produce the purity of silicon required for most semiconductor applications in the necessary quantities, a different approach is used (12). Metallurgical-grade silicon is first converted to a silicon compound that can be easily purified. After extensive purification, that compound is reduced to silicon by using a high purity reducing agent. Silicon tetrachloride, SiCl₄; trichlorosilane, SiHCl₃; and silane, SiH₄, are the silicon compounds most often used. In the earliest commercial process, which was introduced by DuPont, extensively distilled silicon tetrachloride was reduced by high purity zinc to give Si and ZnCl₂. Reduction was carried out in a fused silica tube heated in an electric furnace to about 950°C. The silicon produced was generally in the form of polycrystalline needles from 2.5–5 cm long. Difficulties in handling molten zinc, the clogging of exit lines by ZnCl₂, and the poor packing efficiency of the needles, led to the investigation of many other processes over the years.

The purification method that has become a near-standard is the Siemens process, where hydrogen reduces SiCl₄ or SiHCl₃ on the surface of a resistance-heated (to about 1150°C) high purity silicon rod. The rod is usually U-shaped to reduce the height of the furnace. The result is a silicon ingot several cm in diameter and >2 m long. It is tempting to write the silicon tetrachloride–hydrogen reaction as



but in fact there are many other reaction products. Depending on the temperature and concentrations, many compounds coexist in the Si–Cl–H system. These include SiCl₂, SiH₂Cl₂, SiHCl₃, and a series of long-chain polymeric materials, some of which are explosive. In order to improve process efficiency, silicon manufacturing plants have provisions for reclaiming unreacted feed gases and many of the reaction products. A flow sheet for a facility that starts with metallurgical-grade silicon is shown in Figure 3. In the case illustrated, the fluidized bed is adjusted to produce primarily trichlorosilane, because trichlorosilane is one of the easiest silicon compounds to purify and reduce efficiently. Rather than starting with metallurgical-grade silicon as in Figure 3 and manufacturing the intermediate SiCl₄ or SiHCl₃, some high purity silicon manufacturers take advantage of the large-scale operations of silicone manufacturers and buy SiCl₄ or SiHCl₃ from them. In 1995, whereas the Siemens process was used by the majority of manufacturers, at least two other manufacturers used silane for reduction. One of those latter used a fluidized bed rather than heated rods in the reduction process.

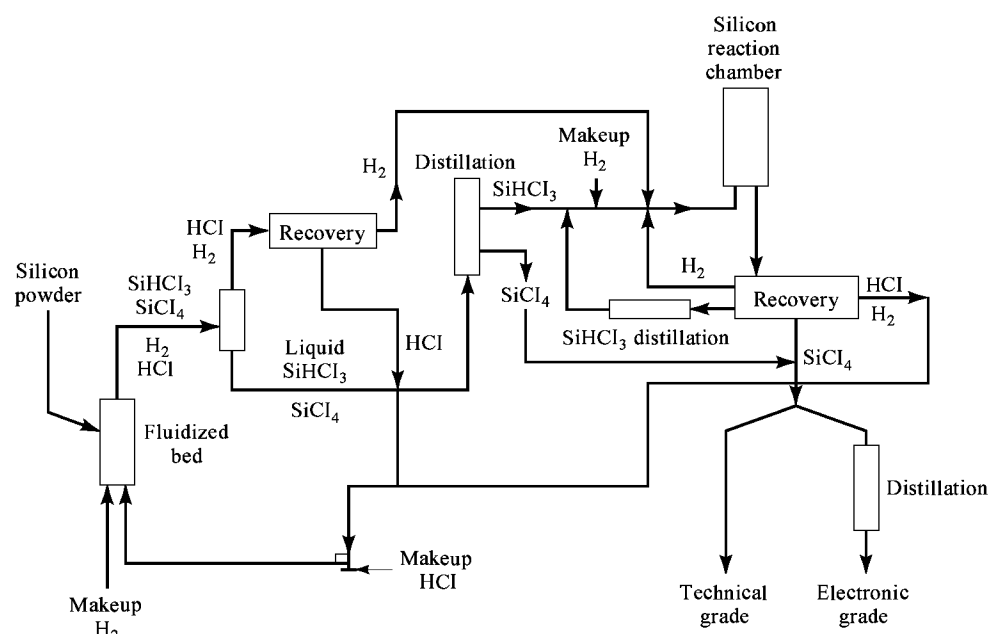


Fig. 3. Flow sheet for silicon manufacturing.

Crystal Growth. Most high purity silicon is used in the manufacture of semiconductor devices, and most such devices are made from single-crystal silicon. Thus the process of growing single crystals of silicon has been studied extensively since the 1950s. A number of processes are available, but the most common is the pulling technique first described by Teal and Little in 1950 (13). This is an adaptation of a much earlier crystal growing method for studying the speed of crystallization of metals (14), called the Czochralski method. The pulling process is shown schematically in Figure 4. Silicon, surrounded by an inert gas atmosphere, is first melted in a fused silica container. Then, the bottom end of a single-crystal silicon seed, capable of being both rotated and pulled up from the melt, is dipped into the melt and slowly withdrawn as silicon freezes on it. The freezing (growth) rate, normally on the order of cm/h, is controlled by a combination of melt temperature, radiation from the growing crystal, and conductive heat losses through the seed. Crystals as small as a few mm in diameter and as large as 300 mm in diameter have been grown by this process. Lengths vary according to equipment design and crystal diameter, but may be as much as >2 m for 100-mm diameter crystals and as little as 50–60 cm for 300-mm diameter crystals.

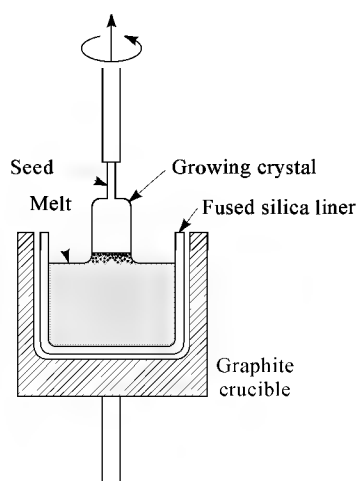


Fig. 4. Crystal pulling from the melt.

The principal difficulty with Czochralski silicon crystal growing is that it is difficult to maintain a container of reactive molten silicon free of contamination for the several hours required to grow a crystal. The contaminants arise from the fused silica container, which is slowly dissolved by the molten silicon, from impurities in the inert atmosphere gas, and from the vaporization of hot portions of the furnace. For applications requiring the highest purity, float zone crystal growing, which uses no crucible and physically uses the same equipment as was shown in Figure 2, is sometimes used instead of crystal pulling. A polycrystalline rod with a single-crystal seed at one end is held vertically, and the heated molten zone is caused to traverse the length of the rod, starting at the end that contains the seed. The molten zone never contacts anything but silicon. The chamber size, and hence the volume of inert gas, can be kept comparatively small; the chamber walls can be kept relatively cool.

Most semiconductor applications require thin, flat sections of crystal having damage-free surfaces for subsequent processing. Both the Czochralski and float zone methods produce long rods from which slices must be cut and subsequently lapped and polished. In an attempt to circumvent these steps, various modifications of melt growth have been proposed in order to grow single-crystal ribbons directly. The work started in the early 1960s concentrated on dendrites and on growth in a $\langle 211 \rangle$ direction through a shaped orifice (15). Neither the dendrites nor the dendrite webs that followed were of sufficient quality for most uses, and shaped crystals could not be grown thin enough to be of practical use. In the mid-1970s, the realization that reasonably efficient solar cells could be made from poorer quality silicon material than was usable for transistors and integrated circuits, combined with a renewed search for low cost cells, led to renewed studies of sheet growth (see SOLAR ENERGY). One of the more promising approaches has been edge-defined growth, where a silicon carbide die having a cross section equal to that of the desired crystal is used. The lower end of the die is immersed in molten silicon in such a way that surface tension can draw molten silicon to the top of the die. The crystal is then pulled from that small reservoir and can grow no larger in cross section than the silicon supply at the top of the die.

Single-crystal silicon can also be grown from various fluxes and by a combination of electrolysis and fluxes at temperatures well below the melting point of pure silicon (16). The main disadvantages are the inclusion of the flux in the crystal and the poor crystal quality. Potential advantages are a decrease in growth temperature and purification during electrolysis.

As of the mid-1990s vapor-phase growth at temperatures well below the melting point is used widely for adding thin single-crystal layers to single-crystal slices. This process is referred to as epitaxy. Vapor-phase growth allows thin layers of high resistivity to be overgrown on low resistivity slices. Moreover, doping impurities in the vapor-grown portion can be rapidly changed, allowing the sequential growth of thin layers having different resistivities.

The method of transport of silicon to the growing interface may be by direct evaporation, ie, molecular-beam epitaxy, or by chemical means (5). Using chemical transport, any of the reactions discussed for manufacturing semiconductor-grade silicon may be used. The thermal decomposition of silane and the hydrogen reduction of principally SiCl₄, SiHCl₃, or SiH₂Cl₂ are used. Deposition rates, which depend on a deposition temperature of typically ca 1000–1100°C, are generally from tenths to a few μm min. Occasionally, polycrystalline or amorphous silicon layers, often on nonsilicon surfaces such as glass or SiO₂, are required, in which case lower deposition temperatures are used.

For adding doping impurities during vapor-phase growth, a gaseous or easily vaporizable liquid compound is metered, added to the silicon source gas stream, and reduced along with the silicon compound. Typical examples are diborane, B₂H₃; phosphine, PH₃; and boron tribromide, BBr₃.

Physical Properties

Unlike structural materials where properties are generally given for polycrystalline samples, many of the properties of the high purity silicon required for the manufacture of semiconductor devices are measured on single-crystal samples. For this reason, even though silicon was isolated in the early 1800s, most of the values for pure silicon have been measured since the introduction of single-crystal silicon growing techniques in the 1950s.

For cubic crystals, which include silicon, properties described by other than a zero- or a second-rank tensor are anisotropic (17). Thus, in principle, whether or not a particular property is anisotropic can be predicted. There are some properties, however, for which the tensor rank is not known. In addition, in very thin crystal sections, the crystal may have two-dimensional characteristics and exhibit a different symmetry from the bulk, three-dimensional crystal (18). Table 4 is a listing of various isotropic and anisotropic silicon properties. Table 5 gives values for the more common physical properties and for some of the thermodynamic properties. Figure 5 shows some thermal properties.

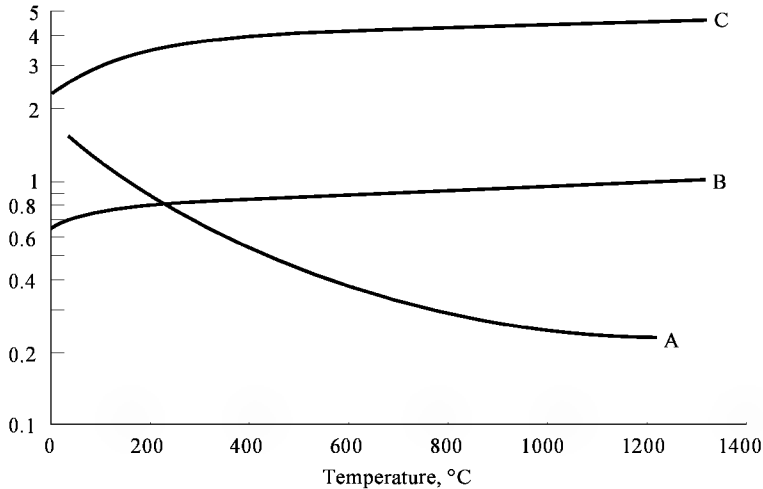


Fig. 5. Thermal properties of silicon (28). A represents thermal conductivity in W (cm·K); B, specific heat in J (g·K) (to convert J to cal, divide by 4.184); and C, thermal expansion coefficient $\times 10^6$ K⁻¹.

Table 4. Isotropic and Anisotropic Silicon Properties

Property	Tensor rank	Anisotropic
density	0	no
heat capacity	0	no
pyroelectric coefficient	1	yes
thermal expansion	2	no
thermal conductivity	2	no
electrical conductivity	2	no
electrical mobility	2	no
diffusion coefficient	2	no
piezoelectric coefficient	3	yes
elastic constants	4	yes
piezoresistance coefficient	4	yes
hardness		yes
breaking strength		yes
chemical etch rate ^a		yes
oxidation rate		yes
ion implant depth		yes
crystal growth rate		yes

^a For some etchants.

Table 5. Properties of Silicon

Property	Value	Reference
density, at 25°C, g cm ³	2.329	19
atomic density ^a , atoms cm ³	5 × 10 ²²	
hardness		20
Mohs'	6.5	
Knoop	950–1150	
elastic constants, GPa ^b		21
C ₁₁	165.7	
C ₁₂	63.9	

C ₄₄	79.57	
Young's modulus in GPa ^b	170 < 110 > direction	22
bulk modulus ^a , GPa ^b	100	
fracture stress, GPa ^b	2.8 ^c	23
melting point, °C	1414	24
volume expansion on freezing, ^o o	9.5	25
boiling point, °C	2355	24
debye temperature, K	645	26
critical temperature, °C	4886	27
critical volume, cm ³ mol	232.6	27
critical pressure, MPa ^d	53.6	27
heat of fusion, kJ mol ^e	50.660	28
heat of vaporization, kJ g ^e	16	27
vapor pressure, Pa ^f		29
800°C	1.33 × 10 ⁻⁸	
1000°C	1.33 × 10 ⁻⁵	
1500°C	2.66	
2000°C	80	

^a Calculated value.
^b To convert GPa to dyn cm², multiply by 10¹⁰.
^c Average value.
^d To convert MPa to atm, divide by 0.101.
^e To convert J to cal, divide by 4.184.
^f To convert Pa to mm Hg, multiply by 0.0075.

Molten silicon is not a semiconductor, and has no commercial use, although because of the high heat of fusion, it has been considered as a heat storage medium. The liquid (molten) silicon properties summarized in Table 6 are nevertheless of importance because these affect single-crystal growth, an operation through which essentially all semiconductor-grade silicon must pass.

Table 6. Properties of Liquid Silicon^a

Property	At mp	At 1500°C	Reference
thermal conductivity, W (mK)	41.84		27
dynamic viscosity, mPas(cP)	0.88	0.7	27
kinematic viscosity, mm ² s(cSt)	0.347 ^b	0.28 ^b	
surface tension, mN m(dyn cm)	736	720	27
heat capacity, J (kgK) ^c	0.16	6.84	27
density, g cm ³	2.533	2.50	27
electrical resistivity, μΩcm	80	100	30
total optical emissivity	0.33	0.33	31
reflectivity at 633 nm, ^o o	72	70	32

^a Many of these values are extrapolated; see Ref. 5 for more details.
^b Calculated value.
^c To convert J to cal, divide by 4.184.

Electrical Properties

Silicon (33) is a semiconductor and thus the electrical conductivity, σ, is determined by contributions from both electrons and holes, ie,

$$\sigma = q(\mu_n n + \mu_p p)$$

(5)

where *q* = charge on an electron, μ = carrier mobility in cm² (Vs), *n* = density of free electrons (electrons cm³), and *p* = density of free holes. For an absolutely pure semiconductor, *n* = *p* = *n_i*. For silicon, *n_i* is given by equation 6:

$$n_i \sim (1.5 \times 10^{33} T^3 \exp(E_g / kT))^{1/2}$$

(6)

where *T* = temperature in K, *k* is Boltzmann's constant, and *E_g* is the band gap in eV. The value of the indirect band gap in silicon can be approximated by equation 7:

$$E_g = 1.16 + 3 \times 10^{-5} T - 9 \times 10^{-7} T^2 + 10^{-9} T^3 - 4 \times 10^{-13} T^4$$

(7)

More precise coefficients are available (33). At room temperature, *E_g* ≈ 1.12 eV and *n_i* ≈ 1.4 × 10¹⁰ cm³. Both hole and electron mobilities decrease as the number of carriers increase, but near room temperature and for concentrations less than about 10¹⁷ there is little change, and the values are ca 1400cm² (Vs) for electrons and ca 475cm² (Vs) for holes. These numbers give a calculated electrical resistivity, the reciprocal of conductivity, for pure silicon of ca 230,000 Ωcm. As can be seen from equation 6, the carrier concentration *n_i* increases exponentially with temperature, and at 700°C the resistivity has dropped to ca 0.1 Ωcm.

Instead of depending on the thermally generated carriers just described (intrinsic conduction), it is also possible to deliberately incorporate various impurity atoms into the silicon lattice that ionize at relatively low temperatures and provide either free holes or electrons. In particular, Group 13 (IIIA) elements (*n*-type dopants) supply electrons and Group 15 (VA) elements (*p*-type dopants) supply holes. Over the normal doping range, one impurity atom supplies one hole or one electron. Of these elements, boron (*p*-type), and phosphorus, arsenic, and antimony (*n*-type) are most commonly used. When

enough impurity atoms are present for holes or electrons from these impurities to overshadow the thermally generated carriers from the silicon itself, conduction is referred to as extrinsic. Impurity concentration is commonly expressed in atoms of impurity per cm^3 of host material. When used as a semiconductor material, purity is sometimes described by the material resistivity, even though resistivity depends on the excess concentration of one doping type impurity over the other ($1000\ \Omega\text{cm}$ silicon is considered very pure; $0.001\ \Omega\text{cm}$ is very impure). Because of the solubility limit of most impurities in silicon, $0.001\ \Omega\text{cm}$ is about as low a resistivity as is observed. The impurity concentration required in a silicon single crystal to be used for semiconductor device manufacture depends on the kind of devices to be made and may range from about 10^{14} to 10^{17} atoms cm^{-3} and give resistivities of from ca 100 to $0.1\ \Omega\text{cm}$.

Carriers can also be generated in a semiconductor by the absorption of light or injected into the semiconductor from a p - n or Schottky junction. In either case, as soon as the source is removed the density of those excess carriers begins to decrease exponentially with time. The time it takes for the density to be reduced to $1/e$ of the original value is defined as the carrier lifetime, τ . For silicon, τ is typically in the microsecond range.

The general behavior of resistivity vs temperature is shown in Figure 6. In the region where extrinsic conduction dominates, resistivity increases with temperature because the carrier mobility decreases with temperature and the carrier concentration remains constant. At some point, however, the temperature becomes high enough for the thermally generated carriers to dominate, and the resistivity drops as more and more carriers are generated (the exponential increase in n_i of eq. 6). The relation between impurity concentration and resistivity at room temperature is shown in Figure 7 (34). The difference between n - and p -type curves arises because of the difference between hole and electron mobilities.

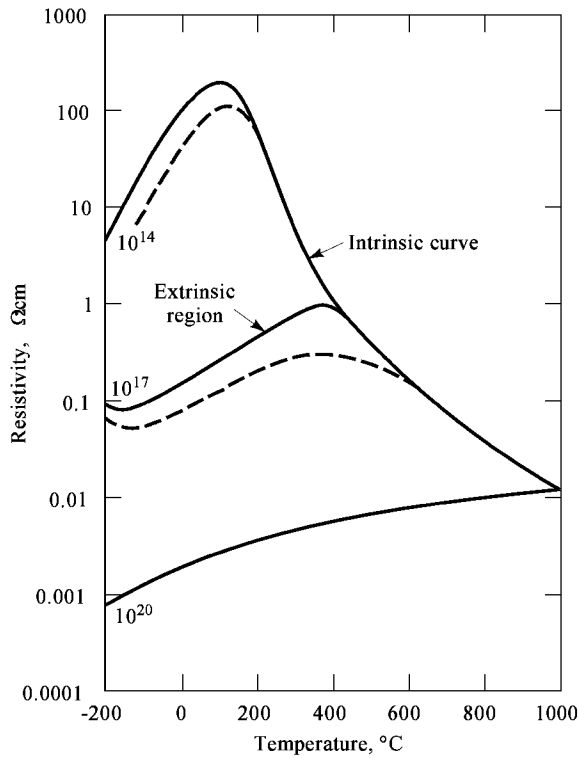


Fig. 6. Calculated silicon resistivity vs temperature for the impurity (doping) levels shown, where (—) is p -type, (---), n -type. Left of the peaks is the extrinsic region where the resistivity is determined by the doping level. To the right of the peaks is the intrinsic region, where the resistivity is determined by thermally generated carriers.

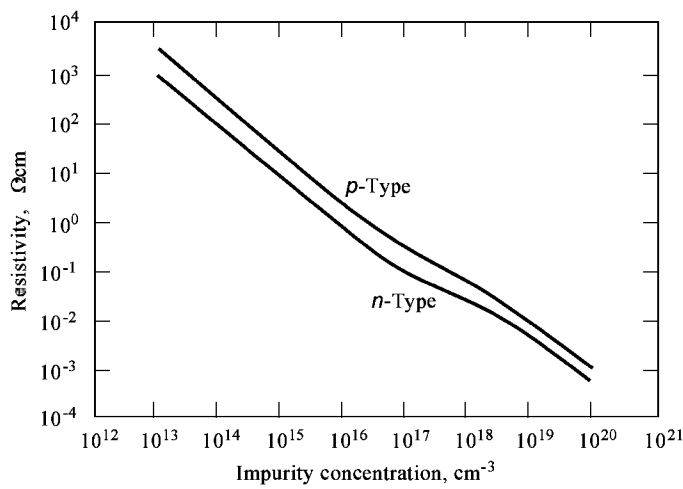


Fig. 7. Impurity concentration vs resistivity at room temperature (34). These curves are for phosphorus (n -type) and boron (p -type). Other Group 13 (IIIA) and 15 (VA) impurities behave similarly.

Courtesy of the Bell System Technical Journal (1960).

In bulk material, the resistivity is independent of crystal orientation because silicon is cubic. However, if the carriers are constrained to travel in a very thin sheet, eg, in an inversion layer, the mobility, and thus the resistivity, become anisotropic (18). Mobility is also sensitive to both hydrostatic pressure and uniaxial tension and compression, which gives rise to a substantial piezoresistive effect. Because of crystal symmetry, however, there is no piezoelectric effect. The resistivity gradually decreases as hydrostatic pressure is increased, and then abruptly drops several orders of magnitude at ca 11 GPa (160,000 psi), where a phase transformation occurs and silicon becomes a metal (35). The longitudinal piezoresistive coefficient varies with the direction of stress, the impurity concentration, and the temperature. At about 25°C, given stress in a $\langle 100 \rangle$ direction and resistivities of a few hundredths of an Ωcm , the coefficient values are $500\text{--}600\ \text{m}^2/\text{N}$ ($50\text{--}60\ \text{cm}^2/\text{dyn}$).

Radiation Effects

Gamma radiation produces free carriers much as does visible light (36). High energy protons and electrons produce defects that reduce minority carrier lifetime according to equation 8:

$$\frac{1}{\tau_f} = \frac{1}{\tau_0} + k\phi \tag{8}$$

where τ_f is the lifetime after irradiation with a fluence ϕ , τ_0 is the original lifetime, and k is a radiation damage constant. Neutrons produce deep-level defects that not only degrade lifetime as in equation 8, but also increase resistivity through the removal of carriers, as described by equation 9.

$$N = N_0 - K\phi \tag{9}$$

where N_0 is the initial carrier density, N the density after irradiation, and K a different radiation damage constant. All of the damage-constant values are dependent on the kind and energy of the irradiation, the kinds of impurities in the silicon, and the temperature of the silicon during irradiation. Moreover, many of the defects anneal out at higher temperatures, thereby reducing the effects of irradiation.

Neutrons also transform some of the silicon atoms to phosphorus through the following reaction:



This reaction is occasionally used for doping crystals uniformly after they have been grown. The process is called transmutation doping (37).

Optical Properties

The optical transmissivity of silicon is shown in Figure 8a. The general characteristic, shared as well by other semiconductors, is that at wavelengths shorter than some critical value defined by the band gap energy, optical absorption generates free carriers and the transmissivity is very low. Figure 8b shows the absorption coefficient for the wavelength range from 0.5–1 μm . The absorption coefficient changes several orders of magnitude over that wavelength range. This change is not nearly as abrupt as it is in direct band gap materials such as gallium arsenide, GaAs. Also, because Si is an indirect band gap material, it cannot be used for conventional light-emitting diodes (see LIGHT GENERATION, LIGHT EMITTING DIODES). The free-carrier generation, shared by both direct and indirect band gap materials, leads to the production of electrical current in photovoltaic devices such as solar cells. The closer the wavelength band of maximum absorption matches the emission peak of the light source, the greater the theoretical maximum efficiency of the cell. For silicon, the match between it and the sun's spectrum is quite good.

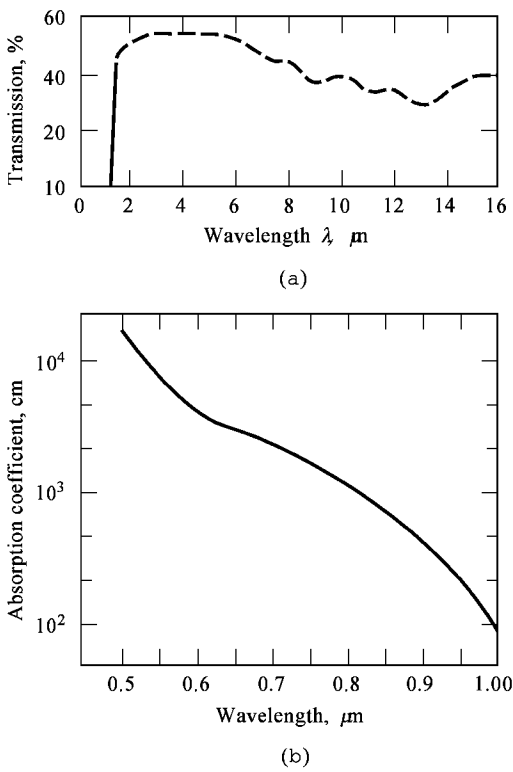


Fig. 8. Optical properties of silicon: (a) transmissivity vs wavelength, (b) shortwavelength absorption coefficient where the transmissivity increases sharply with wavelength.

For wavelengths having photon energies less than that corresponding to the band gap, the transmissivity is substantial, but as wavelengths continue to increase, both impurity absorption bands and the gradually increasing free-carrier absorption reduce transmissivity. There are also a number of lattice absorption bands, the most pronounced of which is at 16.13 nm. Silicon is often used as an infrared optical element because of the good infrared transmissivity in the 1–8- μm region. Owing to the high index of refraction, shown in Figure 9, reflection losses are about 30% per reflecting surface. However, by using antireflection coatings such as silicon monoxide, SiO, those losses can be minimized. Silicon surfaces take a good optical finish. The sawing, grinding, and polishing operations are similar to those for glass.

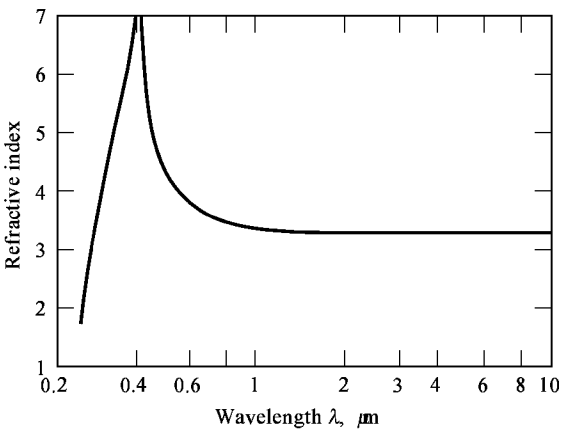


Fig. 9. Refractive index of silicon at room temperature (38).

Economic Aspects

The 1995 world production of high purity silicon was estimated at about 13,000 metric tons. The average selling price was about \$48.00 /kg. Principal suppliers are listed in Table 7. This quantity of silicon supported a silicon device market, ie, primarily diodes, transistors, and integrated circuits, of about \$150 × 10⁹.

Table 7. High Purity Polysilicon Producers in 1995

Producer	Plant location ^a	Quantity ^b , t
Hemlock	United States	3000
Wacker Chemie	Germany	2700
Tokuyama Soda	Japan	1600
Advanced Silicon Materials (Komatsu)	United States	1200
Hi Silicon (Mitsubishi)	Japan	1000
Albermarle (MEMC H&S)	United States	1000
MEMC (H&S)	Italy	600
Sumitoma	Japan	500
	CIS	1100

^a China also produces a small quantity.
^b Values are estimated.

Health and Safety Factors

Elemental silicon is quite inert. In air, silicon is only classified as a nuisance particulate (39) having a threshold limit value (TLV) of 10 mg /m³. However, SiO₂ dust, particularly that of crystalline quartz, is considered hazardous. Many silicon compounds are poisonous, and some, such as the silanes, are highly explosive as well. Compounds such as SiCl₄, SiHCl₃, and SiBr₄ decompose in the presence of moisture, forming SiO₂ and the corresponding acid. If such compounds are breathed, the acid-forming reaction takes place in the moist respiratory tract and the result is similar to that of directly breathing the acid fumes.

BIBLIOGRAPHY

"Silicon (Pure)" in *ECT* 2nd ed., Vol. 18, pp. 111–125, by W. R. Runyan, Texas Instruments, Inc.; "Pure Silicon" under "Silicon and Silicon Alloys", in *ECT* 3rd ed., Vol. 20, pp. 826–845, by W. Runyan, Texas Instruments, Inc.

1. D. R. Lide, ed., *Handbook of Chemistry and Physics*, 76th ed., CRC Press, Boca Raton, Fla., 1995.
2. J. Zhu Hu and co-workers, *Phys. Rev. B*, **34**, 4679–4684 (1986).
3. A. S. Berezhnoi, *Silicon and its Binary Systems*, Consultants Bureau, New York, 1960; J. W. Mellor, *A Comprehensive Treatise on Inorganic and Theoretical Chemistry*, Vol. IV, Longmans, Green & Co., Inc., New York, 1957; M. C. Sneed and R. C. Brasted, eds., *Comprehensive Inorganic Chemistry*, Vol. VII; *The Elements and Compounds of Group IVA*, D. Van Nostrand Co., Inc., Princeton, N.J., 1958.
4. B. E. Deal and A. S. Grove, *J. Appl. Phys.* **36**, 3770 (1965).
5. W. R. Runyan and K. E. Bean, *Semiconductor Integrated Circuit Processing Technology*, Addison-Wesley, Reading, Mass., 1990. See discussion and references.
6. B. Swartz and H. Robbins, *J. Electrochem. Soc.* **123**, 1903 (1976), and preceeding papers in the series.
7. K. E. Peterson, *Proc. IEEE*, **70**, 420 (1982).
8. N. Howell Furman, ed., *Standard Methods of Chemical Analysis*, D. Van Nostrand Co., Inc., Princeton, N.J., 1962.
9. H. C. Torrey and C. A. Whitmer, eds., *Crystal Rectifiers*, McGraw-Hill Book Co., Inc., New York, 1948.
10. W. G. Pfann, *Zone Melting*, 2nd ed., John Wiley & Sons, Inc., New York, 1966.
11. W. Zulehner and D. Huber, *Crystals 8*, Springer-Verlag, Berlin, 1982.
12. L. P. Hunt, in W. C. O'Mara, R. B. Herring, and L. P. Hunt, eds., *Handbook of Semiconductor Silicon Technology*, Noyes Publications, Park Ridge, N.J., 1990, pp. 1–32; L. C. Rogers, *Ibid.*, pp. 33–93.
13. G. K. Teal and J. B. Little, *Phys. Rev.* **78**, 647 (1950); for a review of silicon crystal growing, see W. Zulehner and D. Huber, *Crystals 8*, Springer-Verlag, Berlin, 1982.
14. J. Czochralski, *Z. Physik. Chem.* **92**, 219 (1918).
15. R. O. Grubel, ed., *Metallurgy of Elemental and Compound Semiconductors*, Interscience Publishers, New York, 1961. Discusses early work on semiconductor dendrites and other methods of growing shaped crystals. The special issue of *J. Cryst. Growth* (Sept. 1980) is devoted to shaped crystal growth.

16. D. Elwell, *J. Cryst. Growth* **52**, 741 (1981).

17. J. F. Nye, *Physical Properties of Crystals*, Oxford University Press, Fairlawn, N.J., 1960.

18. D. Coleman, R. T. Bate, and J. P. Mize, *J. Appl. Phys.* **39**, 1923 (1968).

19. I. Hennis, *J. Res. Nat. Bur. Stds.* **68A**, 529–553 (1964).

20. A. A. Giardini, *Am. Mineralogist* **43**, 957–969 (1958).

21. H. J. McSkimin, *J. Appl. Phys.* **24**, 988–997 (1953).

22. J. J. Wortman and R. A. Evans, *J. Appl. Phys.* **36**, 153–156 (1965).

23. S. M. Hu, *J. Appl. Phys.* **53**, 3576–3580 (1982).

24. P. D. Desai, *J. Phys. Chem. Ref. Data* **15**, 967–983 (1986).

25. R. A. Logan and W. L. Bond, *J. Appl. Phys.* **30**, 322 (1959).

26. G. A. Slack and S. F. Bartram, *J. Appl. Phys.* **46**, 89–98 (1975).

27. C. L. Yaws and co-workers, *Solid State Tech.* **24**, **1**, 87–92 (Jan. 1981).

28. *Properties of Silicon*, EMIS Datareview Series No. 4, Inspec, London, 1988.

29. R. E. Honig, *RCA Review*, **23**, 567–586 (1962).

30. V. M. Glazov and co-workers, *Liquid Semiconductors*, Plenum Press, Inc., New York, 1969.

31. S. N. Rea, *Texas Instruments Incorporated Monthly Technical Progress Report*, No. 03-76-31, ERDA JPL 954475-76-II, Texas Instruments, Dallas, May 1976.

32. M. O. Lampert and co-workers, *J. Appl. Phys.* **52**, 4975 (1981).

33. M. Bullis, "Silicon Material Properties", in Ref. 12, pp. 347–450.

34. J. C. Irvin, *Bell System Tech. J.* **41**, 387 (1962).

35. S. Minomura and H. G. Drickamer, *J. Phys. Chem. Solids* **23**, 451 (1962).

36. N. J. Rudie, *Principles and Techniques of Radiation Hardening*, 2nd ed., Vol. 2, Western Periodicals Co., North Hollywood, Calif., 1980.

37. J. Guldberg, ed., *Neutron-Transmutation-Doped Silicon, Proceedings of the Third International Conference on Transmutation Doping of Silicon, Copenhagen, Denmark*, Plenum Press, Inc., New York, 1981.

38. H. R. Philipp and E. A. Taft, *Phys. Rev.* **120**, 37 (1960).

39. *Threshold Limit Values for Chemical Substances and Physical Agents in the Work Room Environment with Intended Changes for 1980*, American Conference of Governmental Industrial Hygienists, 1986.

W. R. Runyan
Texas Instruments Incorporated

CHEMICAL AND METALLURGICAL

Silicon, a low density chemical element having nonmetallic characteristics, is the second, after oxygen (50.5^o %), most abundant element in the lithosphere. Silicon occurs naturally in the form of oxides and silicates and constitutes over 25^o % of the earth's crust (see SILICA).

The commercial production of silicon in the form of binary and ternary alloys began early in the twentieth century with the development of electric-arc and blast furnaces and the subsequent rise in iron (qv) and steel (qv) production (1). The most important and most widely used method for making silicon and silicon alloys is by the reduction of oxides or silicates using carbon (qv) in an electric arc furnace. Primary uses of silicon having a purity of greater than 98^o % are in the chemical, aluminum, and electronics markets (for higher purity silicon, see SILICON AND SILICON ALLOYS, PURE SILICON).

Silicon is soluble in aluminum in the solid state to a maximum of 1.62 wt ^o % at 577°C (2). It is soluble in silver, gold, and zinc at temperatures above their melting points. Phase diagrams of systems containing silicides are available (2,3).

More than half of the elements in the Periodic Table react with silicon to form one or more silicides. The refractory metal and noble metal silicides are used in the electronics industry. Silicon and ferrosilicon alloys have a wide range of applications in the iron and steel industries where they are used as inoculants to give significantly improved mechanical properties. Ferrosilicon alloys are also used as deoxidizers and as an economical source of silicon for steel and iron.

Silicon

Production. Silicon is typically produced in a three-electrode, a-c submerged electric arc furnace by the carbothermic reduction of silicon dioxide (quartz) with carbonaceous reducing agents. The reductants consist of a mixture of coal (qv), charcoal, petroleum coke, and wood chips. Petroleum coke, if used, accounts for less than 10^o % of the total carbon requirements. Low ash bituminous coal, having a fixed carbon content of 55–70^o % and ash content of <4%, provides a majority of the required carbon. Typical carbon contribution is 65^o %. Charcoal, as a reductant, is highly reactive and varies in fixed carbon from 70–92^o %. Wood chips are added to the reductant mix to increase the raw material mix porosity, which improves the SiO (g) to solid carbon reaction. Silica is added to the furnace in the form of quartz, quartzite, or gravel. The key quartz requirements are friability and thermal stability. Depending on the desired silicon quality, the total oxide impurities in quartz may vary from 0.5–1^o %.

The overall reaction for the production of silicon is expressed as follows:



Typical raw material mix to produce one metric ton of silicon consists of 2500–3000 kg quartz, 1200–1400 kg of low ashcoal and or charcoal, and 1500–3000 kg wood chips. From 11 to 14 MWh of electrical power, and prebaked carbon electrodes of 90–140 kg, are consumed.

The production of silicon takes place via two key intermediates, silicon monoxide, SiO, and silicon carbide, SiC. In the high temperature zone of the furnace below and around the electrode, SiC reacts with SiO₂ to produce silicon, SiO, and carbon monoxide, CO, according to equation 2. As the gaseous SiO and CO ascend through the charge bed, the SiO reacts with the carbon reductants to produce silicon carbide according to equation 3. The carbon and silica in the charge bed can also react directly to produce SiC according to equation 4. SiO can also be produced by an equimolar reaction of SiO₂ with carbon (eq. 5).



The silicon carbide, produced in the middle zone of the furnace, is pushed down by stoking and reacts with SiO₂ to produce silicon. The silicon produced is periodically or continuously tapped. Rising SiO reacts with the carbon in the bed and is converted to SiC and the process repeats itself.

A typical 20-MW, a-c furnace is fitted with three 45-in. (114.3-cm) prebaked amorphous carbon electrodes equilaterally spaced, operating on a three-phase delta connection. The spacing of the electrodes is designed to provide a single reaction zone between the three electrodes. The furnace is rotated to give one revolution in two to four days or it may be oscillated only. Rotation of the furnace relative to the electrodes minimizes silicon carbide buildup in the furnace.

The open-top furnace is fitted with a hood to collect the silicon monoxide and carbon monoxide emerging from the furnace. The gases are oxidized by introducing air, which also results in oxidation of the carbon materials on top of the furnace. The off-gas, consisting of silica fume and carbon particles, is passed through radiation pipes and a drop-out box, where the heavy carbon particles are removed. The fume particles are collected in the baghouse and the gases are discharged to the atmosphere.

In a good arc furnace operation, tapped silicon recovery accounts for 80 ± 10% of the inbound silicon in quartz. The remaining 15^o % exits the furnace in the form of SiO, which oxidizes to SiO₂ upon coming in contact with air. The energy required to produce silicon varies from 11–14 kWh /kg Si. Arc furnaces for silicon production are very inefficient from the standpoint of energy utilization. Only 31^o % of the total energy input to the furnace both in the form of electrical energy and chemical energy from the reductants (coal, coke, wood chips) is used for reduction of quartz. The remaining 69^o % of the energy is lost, most of it in the form of by-product off-gas accounting for up to 52.5^o % of the input.

The location of a silicon metal plant is determined by balancing market costs against processing ones. Principal elements in the cost of silicon production, which are site-dependent, are the delivered cost of the raw materials, energy cost, and labor. Typical costs for production of silicon are given in Table 1.

Table 1. Silicon Production Costs in 1996

Operating cost	Cost, \$ /kg Si	Cost contribution, ^o %
----------------	-----------------	-----------------------------------

quartz	0.13	10
carbon	0.22	18
electrodes	0.22	18
electricity	0.26	21
labor	0.17	13
supplies and equipment	0.23	18
working capital	0.02	2
Net operation cost	1.25	100

Silicon production should be a slag-free process. Thus, no slag separation or handling is provided, and the impurities from the raw materials end up in the silicon. Thus, the quality of the silicon is controlled by careful selection of the raw materials. This is especially true for impurities such as Fe and Ti, which are more difficult to remove by refining. A typical analysis of tapped silicon gives 0.2–1.0° Ca, 0.2–0.9° Al, 0.25–0.8° Fe, and <500 ppm for Cr, Mg, Mn, Ni, Ti, V, Cu, Zr, P, and B. The specifications of various grades of silicon for the chemical and aluminum industry are summarized in Table 2.

Table 2. Specifications for Various Grades of Silicon, %

Component	Chemical-grade	Primary aluminum	Secondary aluminum	Metallurgical-grade
Si	99	99	98.5	98.5
Fe	<0.4	<0.35	<0.4	<1.0
Ca	<0.1	<0.5	<0.07	<0.4
Al	<0.2			
Ti	<0.05			
P	<0.01	<0.06		

Refining. In order to produce silicon that meets the requirements of the chemical, ie, silicones, and primary aluminum markets, the silicon produced in the arc furnace requires further purification. The quality of silicon for the chemical silicones industry is critical with respect to the levels of aluminum and calcium present, and the primary aluminum grade of silicon requires low levels of calcium, iron, and phosphorus. The impurity requirements for the secondary aluminum market are not as stringent, so long as the silicon content is >98.5%.

Significant advances have been made toward understanding refining chemistry as well as the practice of refining technology. The concept of gas injection through a bottom plug for refining silicon alloys in a ladle was developed at Tinfos Jernverk A S in Norway during the 1980s (4). Because silicon is refined by mixing air and oxygen with molten silicon, the use of a bottom plug provided significantly improved mass transfer between the impurities in the silicon melt and the oxide slag. This method yields nearly equilibrium levels of calcium and aluminum in silicon. Prior to the development of bottom plug technology, the silicon refining process efficiency was limited by the use of a lance for injection of gases from the top of the ladle. Use of bottom plug technology has commercially demonstrated reduction of calcium levels from 2 to <0.1% and aluminum from 1 to <0.2%.

In order to produce silicon for a given specification with respect to aluminum and calcium, the tapped metal is refined using a predetermined amount of oxygen. The oxygen is introduced to the molten bath as a mixture of oxygen, air, or SiO₂ (sand). The refining efficiency, a function of mass transfer between the silicon melt and the oxide slag, is improved by stirring the bath using the necessary amount of nitrogen, introduced as air. The amount of addition of oxygen as sand depends on tap temperature and heat generated by the oxidation of calcium and aluminum, which is also a function of the tap metal chemistry.

A thorough review of the principles of silicon and ferrosilicon refining has been published (5,6). The equilibrium ternary diagram between the SiO₂–CaO–Al₂O₃ slag and the Si–Al–Ca alloy at 1550°C is shown in Figure 1 (6).

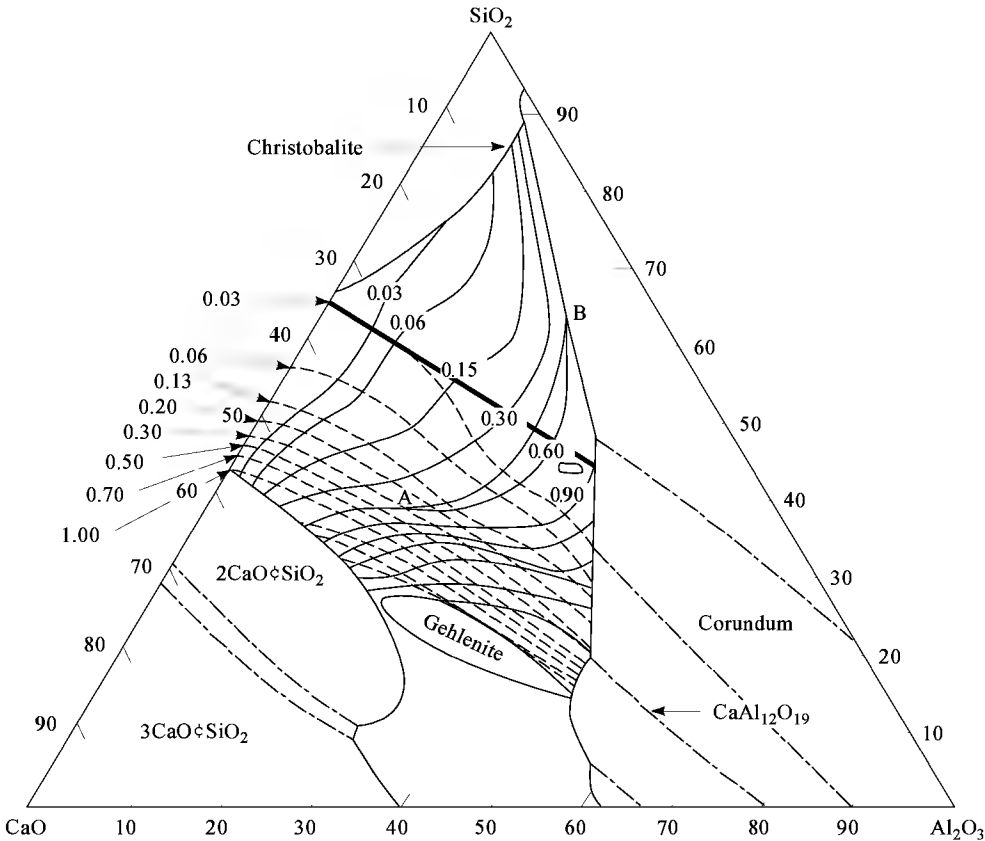


Fig. 1. Equilibrium between SiO₂–CaO–Al₂O₃ slag and Si–Ca–Al alloy at 1550°C (6), where () and (—) represent isoconcentration lines for Ca and Al, respectively, and (—) represent phase transitions in the silicon in equilibrium with the SiO₂–CaO–Al₂O₃ slag. The numbers represent values in wt °. The path designated by A–B represents the change in the slag composition when silicon having 0.6° Al and 0.2° Ca reacts to equilibrium with oxygen;

() corresponds to a slag density of 2.5 g mL.

Technology Developments. There have been significant technology advancements, and these, coupled with the developments in the refining technology, have resulted in significant improvements in product yield and silicon quality. The composite electrode technology is one of the most significant developments put into commercial practice in the early 1990s (7). The improved version of the Sinderberg technology provides iron-free self-baking electrodes which have the potential to reduce electrode cost by about 50% without contributing additional impurities to the product.

Developments in the process computer control technology have been one of the leading factors in improving furnace operation and getting higher silicon recovery. The relationship of resistance and electrode spacing is also a key factor in furnace design optimization. These topics have been thoroughly discussed in the literature (8,9). Other process improvements for improved silicon recovery by carbon theory control via calcium in the tap stream have also been published (10). Calcium addition to a silicon furnace has also been found to improve the operation (11). The direct current closed furnace and the c-c technology for smelting silicon represent an area where a significant amount of effort has been put forth to demonstrate newer technology on a pilot scale (12,13). A parallel effort on smelting silicon and ferrosilicon in a d-c furnace was also evaluated by the Silicon Technology Competitive Cooperative (STCC) (14), the membership of which consists of principal silicon and ferrosilicon producers in the United States.

In the area of silicon casting, water granulation on a commercial scale has been key (15), offering the advantage of uniform distribution of trace impurities within a silicon granule. Some chemical producers have found such silicon to give improved performance (16). Although water-granulated silicon minimizes silicon losses as fines and gives a product that is in demand, it also presents the risk of an occasional explosion, as has been experienced by the producers, when molten silicon comes in contact with water.

Uses. The market for silicon can be broken down into chemical, dominated by silicones; secondary aluminum; primary aluminum, recovery; and electronics. The use of silicon for the silicones market is the fastest growing segment of the silicon market. Silicones are remarkably versatile and can be produced in numerous forms. Examples range from fluids thinner than water to rigid plastics which can be clear or pigmented. These may be elastomeric compounds, heatcured rubber parts, sealants and adhesives, or free-flowing coatings. Silicones can be made to transmit almost no heat or be heat conductive; they can have excellent dielectric characteristics, but be made to conduct electricity. Silicone products are used in a vast number of industries, including aerospace, automotive, chemicals and petrochemicals, construction, electrical, electronics, food processing, paints and coatings, personal household care products, pharmaceuticals, plastics, pressure-sensitive adhesives, textiles, and leather. Some of the products and applications include pressure-sensitive labels for release coatings, drug delivery systems, room-temperature vulcanized (RTV) insulator coatings, hair care products, fabric softeners, foam control agents, optical fiber coating, silicone polyurethane roof system, touch-sensitive computer screens, automobile polishes, fuel-resistant silicones, silicone ink additives, etc. Approximately 25% of the more than 5000 products did not exist before the 1990s. The growth in silicon products has been phenomenal and silicon is the key raw material.

The secondary aluminum market is the largest consumer of silicon. The majority of the applications are in castings for the automotive industry. The most prominently used compositions in all casting processes are those of the aluminum–silicon family. The outstanding effect of silicon in aluminum alloys is the improvement of casting characteristics. Addition of silicon to pure aluminum dramatically improves fluidity, hot tear resistance, and feeding characteristics. It also reduces shrinkage and increases corrosion resistance, hardness, tensile strength, and wear resistance. One of the earliest and still one of the main automotive uses of aluminum alloys is to reduce vehicle weight in components such as radiators, cylinder blocks, cylinder heads, pistons, intake manifolds, crankcases, and compressor, fuel pump, and transmission housings. Another application of aluminum is in rotary air-conditioner compressors. These are made using powder metallurgy (PM) hypereutectic Al–Si alloys (see METALLURGY, POWDER METALLURGY). The wear resistance of new PM aluminum alloys containing 18–19% Si is superior to the conventional die-cast aluminum alloys. Rapidly solidified aluminum alloys using PM to make near-net shapes are also replacing conventional materials in intake valves, connecting rods, and pistons. Aluminum is also being considered for structural components in automobiles with development of successful joining methods.

The third largest silicon consumer is the primary aluminum market. A small (1%) percentage of silicon is added to the virgin aluminum ingot to remove oxygen during casting. Relatively higher purity silicon is required for this application.

The electronics market uses silicon as trichlorosilane, which is decomposed with hydrogen at high temperatures to produce semiconductor-grade silicon (see SILICON COMPOUNDS).

Silicon is also used in the copper (qv) industry for production of silicon bronzes. The addition of silicon improves fluidity, minimizes dross formation, and enhances corrosion resistance and strength.

Besides the chemical industry, silicon is used as a powder in the ceramics (qv) industry for the production of silicon carbide and silicon nitride parts (see ADVANCED CERAMICS). Silicon powder is also used as an explosive for defense applications and in the refractory industry for plasma spraying with other oxide mixtures (see REFRACTORY COATINGS).

Supply and Demand. The demand for silicon metal is driven by consumption in the aluminum and chemical industries. Worldwide consumption of silicon metal grew at an average annualized rate of approximately 5.5% from 1980 to 1995. The growth in the chemical industry grew at a rate of 8.0%, whereas the growth in the aluminum industry was approximately 3.5%. In 1985, Western world consumption of silicon totaled just under 490,000 metric tons, of which approximately one-third was chemicals related, ie, for silicones, electronics, and fumed silica (qv). By 1995, Western silicon demand was 790,000 t. Chemical-related demand more than doubled during this period to approximately 330,000 t, whereas the aluminum-based consumption rose over 35% (17). Western demand is projected to grow at an average rate of 5% into the twenty-first century. By the year 2000, the silicon demand is expected to reach almost 1,000,000 t.

The main impetus for growth in the secondary aluminum market lies in the automotive industry. The weight of aluminum silicon alloys in the automobile was expected to increase from ca 945 kg per vehicle in 1995 to over 1.5 t per vehicle by the year 2000 (18). Growth rate for this market is expected to be around 4% yr through 2000. For the primary aluminum market, which represented 82,000 t of demand in 1995, the growth rate is expected to be about 1% yr through 2000. For the electronics market, which is expected to grow at a rate of 12% yr through 2000, the 1995 demand was estimated at 25,000 t yr.

The demand for silicon in the 1990s has exceeded the installed Western world capacity. The difference has been supplemented by shipments from China, the Ukraine, and Russia. In 1993, Chinese exports reached 117,000 t, whereas exports from Ukraine and Russia were around 40,000 t (19). In 1995, the exports from China increased to 155,000 t, whereas the exports from the CIS (former USSR) countries declined. The silicon metal shipped from these countries has been high in iron and calcium and has been used primarily in the secondary aluminum market.

Economic Aspects. Silicon metal is produced in the United States by American Silicon Technologies (Wenatchee, Washington), Elkem Metals Company (Alloy, West Virginia), Globe Metallurgical Inc. (four sites at Selma, Alabama; Beverly, Ohio; Springfield, Oregon; and Niagara Falls, New York), and SIMCALA (Montgomery, Alabama). In Canada the only company producing silicon metal is SKW Canada Ltd. (Becancour, Quebec). In 1990, the producers of silicon metal in the United States filed for antidumping and countervailing duties against imports of silicon metal from Argentina, Brazil, and China. In 1991, the U.S. Department of Commerce (DOC) and the U.S. International Trade Commission (ITC) assessed the dumping duty for Argentina at 8.65% and China at 139.49%. Duties from Brazilian imports were assessed against individual producers over a range of 87.79–93.2% (20). Silicon metal imports into the United States from Brazil and China dropped substantially following the determinations. In December 1993, DOC and ITC amended the antidumping duty rate on Argentine silicon metal, raising it from 8.65% to 54.67% (21). The antidumping duty against Electrosil, a producer from Brazil, was also reviewed in early 1994 and assessed lower. These antidumping duties against the foreign producers are in force as of this writing (ca 1996).

In a parallel move, antidumping duties against these three countries were also assessed for import of silicon into the European Community. These latter duties are due to expire. The term for antidumping duties in Europe is a period of five years. The impact of antidumping duties is not straightforward, but rather is influenced by factors such as Most Favored Nation Status in the United States and by export of finished products made from materials containing antidumping duties tariff to Third World nations and in the European Community.

The only new silicon capacity since the early 1990s is by the Gulf Ferroalloys Company in Saudi Arabia. This plant is expected to have four furnaces and produce silicon metal, ferrosilicon, silicomanganese, and manganese alloys by late 1996.

The average annual import price of silicon metal based on *Metals Week* U.S. dealer import price is shown in Figure 2 (22). The price of silicon over

The 15 years prior to 1990 was quite stable, except for 1988 when it increased by about 10%. Beginning in 1990, silicon price gradually increased. The steepest rise was in 1996, when in April the price of silicon was quoted at \$1.98/kg, a 35% increase over the 1995 average price.

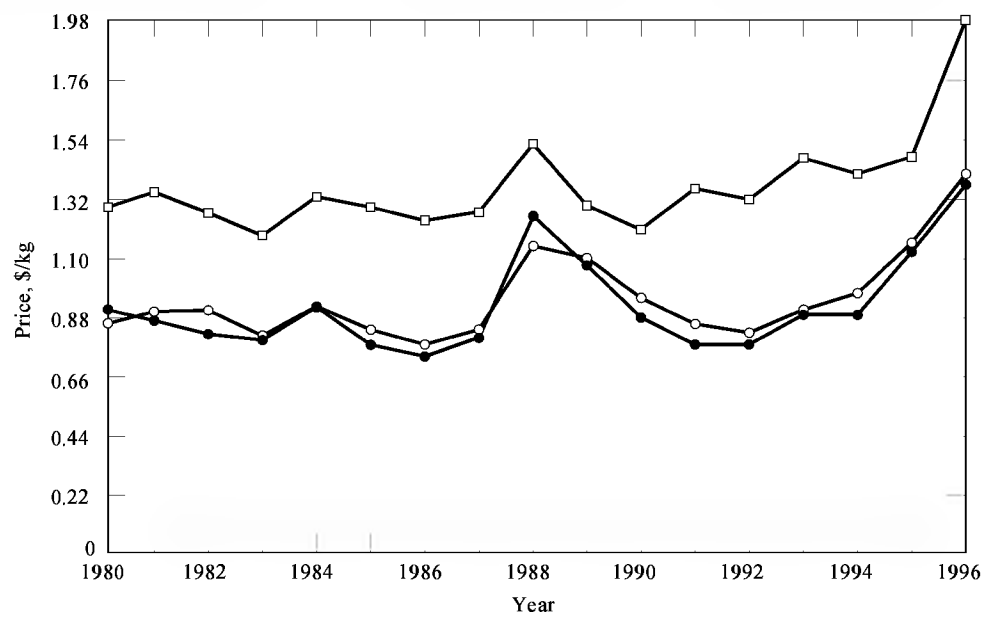


Fig. 2. Silicon product prices where (●) represents silicon metal; (○), 50% FeSi; and (□), 75% FeSi (22).

Silicon Alloys and Silicides

silicides are compounds of silicon and a metal. More than half the elements in the Periodic Table react with silicon to form one or more silicides. Figure 3 gives the known silicides, arranged according to their position in the Periodic Table (23). The silicides of Groups 4–6 (IVA, VA, VIA), the refractory metal silicides, and of Groups 8–10 (VIII), the near-noble metal silicides, are of importance for very large-scale integration (VLSI) electronic applications (see INTEGRATED CIRCUITS). Metal silicides form well-defined crystals with a bright metallic luster. These are usually hard and high melting. Metals that do not form silicides are Al, Sb, As, Bi, Ca, Au, Hg, Ru, Ag, Na, Th, and Zn (2).

1 (IA)	2 (IIA)	3 (IIIA)		4 (IVA)	5 (VA)	6 (VIA)	7 (VIIA)	8	9 (VIII)	10		11 (IB)	12 (IIB)	13 (IIIB)	14 (IVB)	15 (VB)	16 (VIB)	17 (VIIB)	18 (O)
H ₄ Si																			
Li ₁₅ Si ₄ Li ₂ Si														B ₆ Si B ₄ Si B ₃ Si	CSi	N ₄ Si ₃	OSi O ₂ Si	F ₄ Si	
NaSi	Mg ₂ Si														Si	PSi	S ₂ Si	Cl ₄ Si	
KS KS ₆	Ca ₂ Si CaSi CaSi ₂	Sc ₅ Si ₃ ScSi Sc ₂ Si ₃ Sc ₃ Si ₅	Zr ₄ Si Zr ₂ Si Zr ₃ Si ₂ Zr ₆ Si ₃ ZrSi ZrSi ₂	Nb ₄ Si Nb ₅ Si ₃ NbSi ₂	Mo ₃ Si Mo ₅ Si ₃ Mo ₃ Si ₂ MoSi ₂		Ru ₂ Si RuSi Ru ₂ Si ₃	Rh ₂ Si Rh ₅ Si ₃ Rh ₃ Si ₂ RhSi Rh ₂ Si ₃	Pd ₃ Si Pd ₂ Si PdSi	Cu ₃ Si				As ₅ Si AsSi	Se ₂ Si	Br ₄ Si			
RbSi RbSi ₆	SrSi SrSi ₂	Y ₅ Si ₄ Y ₅ Si ₃ YSi YSi ₂	Hf ₆ Si Hf ₅ Si ₂ Hf ₃ Si ₂ HfSi HfSi ₂	Ta _{4,5} Si Ta ₂ Si Ta ₅ Si ₃ TaSi ₂	W ₃ Si W ₅ Si ₃ WSi ₂	Re ₃ Si Re ₅ Si ₃ ReSi ReSi ₂	OsSi OsSi ₂ OsSi ₃	Ir ₃ Si Ir ₂ Si Ir ₃ Si ₂ IrSi IrSi ₃	Pt ₃ Si Pt ₂ Si PtSi					Te ₂ Si TeSi	J ₄ Si				
CaSi CaSi ₃	BaSi BaSi ₂	La ₅ Si ₃ LaSi LaSi ₂																	
			Ce ₃ Si Ce ₂ Si CeSi CeSi ₂	PrSi ₂	NdSi ₂		SmSi ₂		Gd ₃ Si ₅ GdSi ₂		Dy ₃ Si ₅ DySi ₂		Er ₃ Si ₅		YbSi _x	Lu ₂ Si ₅			
			Th ₃ Si ₂ ThSi ThSi ₂		U ₃ Si ₂ USi U ₂ Si ₃ USi ₂ USi ₃	NpSi ₃	PuSi PuSi ₂												

Fig. 3. Silicides of elements in the Periodic Table (23).

Ferrosilicon and Alloys. Ferrosilicon is a grayish ferroalloy consisting mainly of the elements silicon and iron. Ferrosilicons for steelmaking and foundry uses have a silicon content of 14–95 wt %. The Fe–Si phase diagram (Fig. 4) shows the existence of four compounds: Fe_3Si , Fe_5Si_3 , Fe_2Si , and FeSi_2 . Silicon is an essential element in cast iron. The silicon level is often the controlling factor in whether cast iron solidifies as gray iron having its accompanying mechanical properties or as white iron or iron carbide, which has little engineering significance because of its brittle nature.

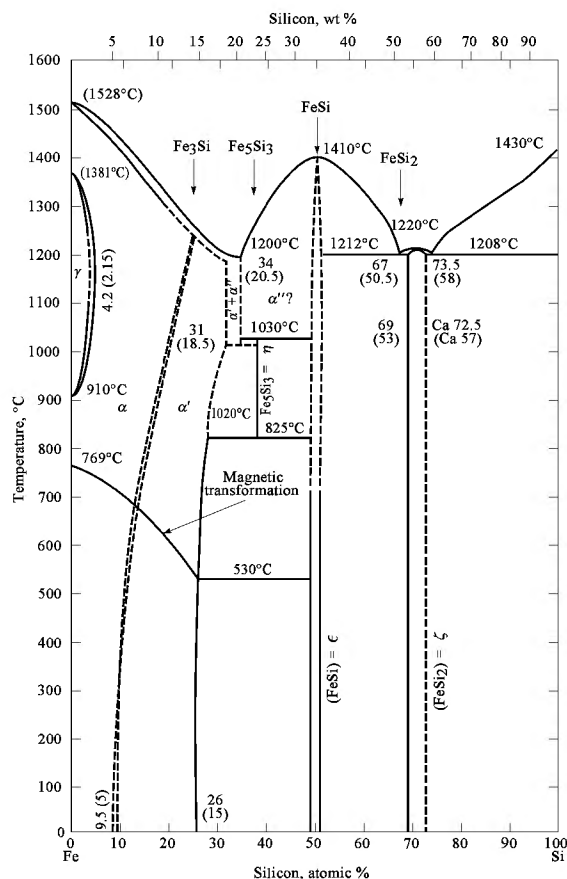


Fig. 4. Fe-Si phase diagram (2). Numbers in parentheses represent composition in wt %.

Courtesy of McGraw-Hill Book Co., Inc.

Ferrosilicon Production. Ferrosilicon is produced in a three-phase submerged arc furnace by the carbothermic reduction of quartz in the presence of high quality iron ore or scrap steel. The smelting process for ferrosilicon is quite similar to silicon, with the exception of the amount of silicon carbide that is formed as an intermediate. As the iron content of ferrosilicon increases, the stability of the silicon carbide phase is known to decrease. In grades of ferrosilicon containing less than 22% Si, silicon carbide is not formed as an intermediate and quartz is directly reduced by carbon in the presence of iron. The presence of iron during the carbothermic reduction of quartz lowers the partial pressure of SiO required for reduction to silicon. Thus the losses of SiO from the furnace top decrease as the iron content of ferrosilicon increases. Because of this thermodynamic consideration, it is typical to have silicon recovery of greater than 90% for 50% and lower grades of ferrosilicon. Another significant difference in ferrosilicon production is the use of Sderberg electrodes, which utilize a carbon paste that is baked *in situ*. Ferrosilicon production also utilizes larger furnaces having a power rating of up to 70 MW.

Ferrosilicon production is a nearly slag-free process. Most of the impurities introduced into the furnace via the raw materials and any other sources are transferred to the product. In order to produce high purity grades of ferrosilicon, the tapped alloy is refined by treating with gas mixtures and slag additions. The principle of ferrosilicon refining is quite similar to that of silicon. Details on the thermochemistry of ferrosilicon refining and practical examples are available (24).

Processes for casting ferrosilicon are quite similar to those used for silicon. During solidification of the standard grade of 75% ferrosilicon, silicon crystallizes initially, followed by solidification of the ϵ -phase. The transformation of the ϵ -phase to stable FeSi_2 is accompanied by an increase in volume, which can lead to disintegration of the alloy in the range of 45–65% silicon. The transformation and thus the disintegration can be suppressed by casting ferrosilicon in shallow molds. Ferrosilicon may be granulated by pouring a controlled molten stream of alloy in a tank of water. The range of resulting granule size is 0.32–1.92 cm (1/8–3/4 in.). Alloys of silicon are of significant commercial importance and are typically produced by post-taphole ladle additions.

Silvery pig iron containing <25% Si is magnetic, which is an advantage in handling. It is used primarily as a furnace block and is added in the form of controlled-weight piglets for initial deoxidization (25,26). It is also furnished in powder form for ore beneficiation (26). The low silicon content of silvery pig iron makes it unsuitable for ladle addition, because the large quantity normally required would cause excessive chilling.

Specialty Silicon Alloy Uses. In the iron and steel industry, silicon alloys or silicides are used for alloying, deoxidizing, and reducing other elements such as manganese, chromium, tungsten, and molybdenum. Silicon in the form of 75% ferrosilicon is used as a reducing agent in magnesium manufacture by the Pidgeon process. Ferrosilicons containing 14–95% silicon are used extensively by the iron and steel industry, and silicon is present in most commercial grades of steel and cast iron. Ternary and quaternary ferroalloys, often proprietary, are also available for the simultaneous addition of silicon and other alloying elements, usually manganese and iron. Other elements added in this way include aluminum, boron, barium, calcium, chromium, magnesium, strontium, titanium, vanadium, zirconium, and the rare earths. These combinations are widely used in the iron foundry and steel industries.

As an effective and economical deoxidizer, silicon is used to refine most grades of carbon and alloy steels. In importance to steelmaking, it is second only to manganese. As an alloying element in steels, it increases tensile strength and elastic limits, improves resistance to corrosion and high temperature oxidation, improves electrical characteristics, and decreases yield point. As a reducing agent, it reduces metal oxides from slag, thereby permitting the desired element, such as chromium, to sink and be recovered as an alloying agent. Silicon in cast iron reduces the stability of iron carbide and promotes the formation of graphitic carbon. As an effective graphitizer in cast iron, silicon softens the iron and improves its fluidity and machinability. For wear resistance, silicon in gray iron ranges from 0.50–1.50%. In ductile iron, silicon ranges from 1.50 to 3.00%. Increased percentages of silicon improve the corrosion and oxidation resistance of gray and ductile cast irons.

Regular Ferrosilicon. Regular 50% ferrosilicon is a widely used, economical source of silicon for both iron and steel. Unlike steel scrap, 50% ferrosilicon provides a known source of high quality iron. This ferrosilicon is used as a deoxidizer and alloying agent in the production of killed and semikilled steels. Killed steels are those that have been made free from bubbling, while molten, by the addition of a deoxidizer. This process minimizes reaction of carbon and oxygen during solidification. The melting point of 50% ferrosilicon at 1220°C is the lowest of the iron–silicon alloys, except for the 1200°C eutectic for silvery pig iron containing about 22% Si (see Fig. 4). Iron foundries add 50% ferrosilicon to the cupola and subsequently to the ladle for the purpose of inoculation. Inoculation may be defined as a change in the physical properties of iron that are not explainable by a change in chemistry. In practice, inoculation consists of withholding from the base iron some portion of the desired silicon and adding it to the ladle. The calcium and aluminum levels in these alloys effectively combine with other elements such as sulfur and oxygen in liquid iron to form oxy–sulfide nucleation sites. The nucleation sites assist in the formation of graphite and allow the iron to solidify without undercooling or the formation of chill or carbides without appreciable effect

on pearlitic stability, increased randomness of graphite flakes and decrease in flake size (decreased section sensitivity), increased ratio of tensile strength to hardness, and improved machinability at all strength levels. Regular 50% ferrosilicon containing 0.04–0.1% boron is used in the production of malleable castings, where small amounts of boron accelerates the annealing cycle.

Regular 75% ferrosilicon is widely used for deoxidizing and alloying additions to iron and steel that do not require tight control of residuals. This ferrosilicon is ideal for ladle additions because its higher silicon content permits smaller additions in order to reach desired silicon levels. The slight exothermic nature of 75% ferrosilicon is also desirable from the standpoint of ladle heat balance. Regular 75% ferrosilicon and 90% ferrosilicon grades are used mainly for high alloy cast steels that require large additions of silicon. Regular 75% ferrosilicon can also be used to reduce the chromium oxide in the slag of stainless steel produced by the argon oxygen decarburization (AOD) process. Because 75% ferrosilicon is not quite as dense as steelmaking slags, additions are best added to the ladle during tapping or refining.

In the production of cast irons, 50 and 75% ferrosilicon can both be used in either cupola charges or electric furnace melting. If density is a consideration in electric furnace melting, 50% ferrosilicon has a slight advantage, because 50% ferrosilicon tends to sink deeper and penetrate the slag layer, owing to its higher density (Table 3). In cupola applications, 75% ferrosilicon is preferred because it dissolves exothermically at a faster rate (Table 3) than 50% ferrosilicon and consequently penetrates further into the melt zone, resulting in slightly higher recoveries.

Table 3. Ferroalloy Density and Thermal Effects on Steel Baths

Alloy	Apparent density, g/mL	Temperature change from adding 0.2% silicon, °C	
silicon	2.3	2.78	
ferrosilicon			
75%	2.9	0.56	
50%	4.9		−4.44
iron	6.9		
steel	7.2		
slag	3.3		

High aluminum grade of 75% ferrosilicon containing 3–5% Al is used for the production of spheroidal iron. In gray iron applications, this inoculant is used to control and minimize the formation of carbides and decrease section sensitivity by refining graphite size and distribution. In ductile iron it is also effective in minimizing carbide formation while increasing nodule count.

High purity 50% ferrosilicon containing <0.1% Al and C is used for production of stainless steel and corded wire for tires, where residual aluminum can cause harmful alumina-type inclusions. These are also useful in continuous cast heats, where control of aluminum is necessary. High purity grades of 50 and 75% ferrosilicon containing low levels of aluminum, calcium, and titanium are used for silicon additions to grain-oriented electrical steels, where low residual aluminum content contributes to the attainment of desired electrical properties, eg, significant reduction of eddy currents.

High purity 75% ferrosilicon containing low aluminum and calcium can be used in continuously cast heats where nozzle blockage is a problem. In iron melting, this alloy is desirable to minimize buildup on refractory faces in the furnace or on stopper rods or nozzles. Low aluminum ferrosilicons can also help reduce hydrogen pin holes in castings poured in green-sand molds.

Ferrosilicon Supply and Demand. Western world production of ferrosilicon amounted to 1,120,000 metric tons of contained silicon in 1993, the equivalent of 1,500,000 t of 75% ferrosilicon. The Western world production has fallen by almost one-fourth since the all-time peak recorded in 1989. The main reason for this drop has been the rise in ferrosilicon exports from China, Eastern Europe, and other CIS countries.

In 1994, there were approximately six to seven dozen companies producing silicon in about 30 different Western countries, together employing several hundred furnaces. Nameplate Western world ferrosilicon capacity was estimated at 1,830,000 t of contained silicon. The effective capacity in 1994 was estimated at 1,450,000 t of contained silicon. There has been a definite shift in ferrosilicon capacity and trends for new plants and expansions are outside of the traditional supply areas. Since 1992, new capacity has been added in South Africa, Venezuela, Iran, and Bhutan. Additional plants near completion in Iran and Saudi Arabia are scheduled to start in 1996. The Western world ferrosilicon supply is expected to increase by about 150,000 t/yr of contained silicon by the year 2000 (27).

Economic Aspects. In June 1992, the U.S. International Trade Commission and the Department of Commerce announced antidumping duties for import of ferrosilicon from China, Kazakhstan, Russia, Ukraine, and Venezuela. The DOC assessed the dumping duty for China at 137.73%; Kazakhstan, Russia, and Ukraine at 104.18%; and Venezuela at 9.55%. Also, Venezuela was assessed a countervailing duty of 22.08%. Duties from Brazilian importers were assessed against six individual producers: Italmagnesio, 88.86%; Rima, 35.95%, Paulista, 35.95%; Ferbasa, 35.95%; CBCC, 15.53%; and Minasligas at 3.46% (28). The European community also exercised antidumping duties at somewhat different rates for import of ferrosilicon against all the above listed countries. South Africa was also added to the list (27). Despite the antidumping duties against import of ferrosilicon into the United States, Elkem Metals Company announced shutdown of its 50-MW ferrosilicon furnace at its plant in Ashtabula, Ohio.

The prices of 50 and 75% grade ferrosilicon are shown in Figure 2. The prices are given on a contained silicon basis (22). The price of ferrosilicon increased significantly in 1988 compared to previous years and gradually declined until 1992. Beginning in 1992, ferrosilicon prices have gradually increased, posting significant gains in 1995 and 1996 owing to increased demand in both 50 and 75% grades of ferrosilicon.

Specialty Silicon Alloys

Magnesium Ferrosilicons. Of the miscellaneous silicon alloys consumed in the United States, magnesium ferrosilicon is the most significant. Magnesium ferrosilicon alloys have a light gray to silvery crystalline appearance and are primarily used in the production of ductile iron. Magnesium ferrosilicon alloys are produced by post-taphole ladle additions of magnesium and at times rare earths to the ferrosilicon production process. In order to improve magnesium recovery and uniformity, the alloys are produced using thin casting techniques in chill molds. The rare-earth-free grades typically contain 44–45% Si, 3–9% Mg, 0.8–1.5% Ca, and 1.25% (max) Al. The cerium-bearing grades contain 0.2–2% Ce in addition. The content of rare earths range from 0.2–2.5%. Rare-earth-containing magnesium ferrosilicons are also available with high calcium contents of up to 3.4%. The presence of calcium slows down the dissolution rate of magnesium ferrosilicon, providing uniform magnesium recoveries.

Magnesium ferrosilicon alloys react vigorously when added to molten iron. As the magnesium vaporizes and cools, it reacts with residual surface tension modifiers such as sulfur and oxygen and greatly increases the surface tension of the molten iron. The dissolved graphite in the molten iron nucleates and grows into a spheroidal shape because of the increased surface tension of the molten iron.

High quality ductile iron can be produced by closely controlling impurities in base irons without adding rare earth. Rare-earth-free grades of magnesium ferrosilicons are used for heavy section ductile iron castings where high quality scrap or pig iron is used in large quantities. Presence of cerium improves nucleation, neutralizes the harmful contaminants in base iron, and reduces chill in this section of iron castings.

Ductile cast iron is made by converting the flakes of graphite in gray iron into tiny balls or spherulites by the addition of one or more elements to the molten metal. Magnesium, calcium, cerium, barium, or other elements produce spherulitic graphite structures (29); magnesium and cerium (rare earth) are commercially important. The addition of 1.50–1.95% nickel to nodular iron significantly increases the strength of the pearlitic matrix. Conversion of graphite flakes into spherulites increases tensile strength and notch hardness of cast iron; subsequent heat treatment produces desirable mechanical properties that cannot be obtained with iron-containing graphite in flakes.

Rare-Earth Silicides. Rare-earth silicides, in the form of a ferroalloy that contain up to 33% rare earths, are used increasingly by the iron and steel industries. Whereas the term silicides is no longer used for alloys of this type, it is still in common usage for these materials. For nodular iron, addition

of rare earths gives spheroidal rather than flaky graphite. Rare earths desulfurize and, combined with silicon, deoxidize and alloy steels. They may be substituted for mischmetal.

Manganese–Silicon Alloys. Manganese–silicon alloys usually contain various amounts of iron and carbon. They are made by carbon reduction of manganese ore or manganese slag in the presence of silica in an electric furnace. The lower silicon grades, referred to as silicomanganese, contain ca 1.5–3% carbon, 18–20% silicon, and 65–68% manganese. The remainder is mainly iron. Silicomanganese is widely used in steelmaking furnaces and ladles for alloying and deoxidization. The use of silicomanganese is often preferred to separate additions of manganese and silicon because it introduces less aluminum, carbon, nitrogen, and phosphorus. It also provides stronger deoxidization and produces higher purity steel. A low carbon grade of silicomanganese containing about 30% silicon and 0.6% carbon (max) is made by a two-stage process. It is used extensively in the production of the 200 and 300 stainless steel series.

Barium- and calcium-bearing manganese silicon is used as an inoculant in gray and ductile iron. The alloy contains 60–65% Si, 9–11% Mn, 4–6% Ba, 1–3% Ca, and 1–1.5% Al. The combination of barium, calcium, and manganese provides excellent chill reduction, improves the graphite structure, and minimizes section sensitivity in castings having thin and thick sections.

The zirconium-bearing manganese silicon, having a composition of 60–65% Si, 5–7% Mn, 5–7% Zr, 0.75–1.25% Al, and 0.6–0.9% Ba and Ca, is also used as an inoculant for gray and ductile iron castings. This alloy minimizes chill tendency while improving the mechanical properties of cast iron. Manganese in the alloy enhances its dissolution rate, strengthens the matrix, and improves tensile strength without impairing machinability. Calcium and barium help reduce chill and increase cell or nodule count. Zirconium acts as an effective nitrogen stabilizer, and the presence of barium and zirconium has a synergistic effect on minimizing the rate of inoculation fade.

Calcium–Silicon. Calcium–silicon and calcium–barium–silicon are made in the submerged-arc electric furnace by carbon reduction of lime, silica rock, and barites. Commercial calcium–silicon contains 28–32% calcium, 60–65% silicon, and 3% iron (max). Barium-bearing alloys contains 16–20% calcium, 9–12% barium, and 53–59% silicon. Calcium can also be added as an alloy containing 10–13% calcium, 14–18% barium, 19–21% aluminum, and 38–40% silicon. These alloys are used to deoxidize and degasify steel. They produce complex calcium silicate inclusions that are minimally harmful to physical properties and prevent the formation of alumina-type inclusions, a principal source of fatigue failure in highly stressed alloy steels. As a sulfide former, they promote random distribution of sulfides, thereby minimizing chain-type inclusions. In cast iron, they are used as an inoculant.

Barium–Silicon. Barium–silicon alloy, containing 70–77% Si, 1.5–2.5% Ba, 1.7–3% Ca, and 0.8–1.5% Al, is used as an inoculant in both gray and ductile iron. The combined effect of barium and calcium provides extremely potent nucleation and is very effective in minimizing carbide precipitation and decreasing section sensitivity by refining graphite size and distribution in gray iron applications. In ductile irons, it is also effective in minimizing carbide formation while increasing nodule count.

Strontium–Silicon. The strontium–silicon alloy contains 30–40% strontium and 45–55% silicon; 0.9% Ca, 1% Al, and 5% Fe. It is made in an electric furnace by carbon reduction of silica- and strontium-bearing ore. In cast irons, this alloy effectively reduces chill. It dissolves rapidly, produces little dross, and minimizes shrinkage defects. This alloy is also used to make a master alloy containing 9–11% Sr, 12–16% Si, and the remainder aluminum. This master alloy is used for additions to aluminum to improve machinability. Strontium is an alternative to sodium in this application. Strontium–silicon can also be used as a strontium source to produce an alloy having either 50 or 75% ferrosilicon and a strontium level in the range of 0.6–1%. This alloy is very effective in controlling graphite structure by providing an excellent nucleating effect and reducing chill in gray iron. It also helps avoid casting shrinkage and pinhole defects.

Titanium–Silicon. The titanium–silicon alloy is made by adding titanium scrap to molten metallurgical silicon or silicon alloys. It is also made by carbon reduction of titanium ore, limestone, and quartz in a submerged arc furnace. Titanium–silicon alloy is an efficient graphitizing inoculant for chill reduction in gray cast iron. It is also a supplementary deoxidizer for wrought and cast steels. The high calcium-based titanium silicon alloy contains 50–60% Si, 9–12% Ti, 5–8% Ca, and 0.9–1.4% Al and is used primarily for steel deoxidization. The high calcium level permits deoxidization to occur with reduced aluminum additions. The titanium readily combines with dissolved nitrogen in liquid steels eliminating subsurface porosity. The low calcium grade containing 50–55% Si, 10–12% Ti, 0.5–1.5% Ca, 0.9–1.2% Al, and 1.2% Ba (max) is used primarily for inoculation of gray iron castings to eliminate nitrogen porosity.

The high barium grade, titanium-bearing, 75% ferrosilicon is used as an inoculant aimed at reducing nitrogen porosity, while simultaneously providing superior inoculating effects for controlling graphite structures and reducing chill. Typical composition consists of 65–70% Si, 8–10% Ti, 5% Mn (max), 2–2.5% Ba, 1–1.5% Ca, and 1.5% Al. The manganese in the alloy produces a mild pearlite stabilizing effect. Pearlitic matrix structures in gray irons enhance tensile properties.

An alloy of titanium containing 40–50% Ti and 45–50% Si is used as an additive in cast iron to shorten the graphite flakes. The effect is to provide a smooth casting surface. The resulting casting is then used to produce glass bottle molds.

Vanadium–Silicon. Vanadium–silicon alloy is made by the reduction of vanadium oxides with silicon in an electric furnace. Application is essentially the same as that of the titanium alloys. Vanadium alloys sometimes offer the most economical way of introducing vanadium into molten steel.

Zirconium–Silicon. Zirconium–silicon alloy is made in an electric furnace by carbon reduction of the oxides. It is used by the iron and steel industries as a deoxidizer and scavenger. Zirconium combines readily with oxygen, nitrogen, and sulfur, forming nonmetallic inclusions that either float out of the molten bath or are rendered minimally harmful. The principal alloy contains a typical level of 35–40% Zr, 49% Si, and 10% Fe. A lower zirconium product containing about 14% Zr, 42% Si, and 42% Fe has also been produced. Zirconium–silicon alloy is mainly added to steel as a grain refiner, but can also be used for oxygen and nitrogen removal. When added to high strength low alloy steels, it serves as a sulfide inclusion modifier.

Zirconium-bearing 75% ferrosilicon is used as an inoculant in both gray and ductile iron. The alloy contains 70–80% Si, 0.4–1.5% Ca, 1–1.8% Al, and 0.25–3.5% Zr. Because of its high dissolution rate, it can be added in the ladle or in-stream inoculant. This alloy gives excellent chill reduction.

Health, Safety, and Environmental Factors. The off-gas from the production of silicon and ferrosilicon alloys contains large amounts of dust having approximately 90% silica fume. This nonhazardous silica fume is amorphous and more of a nuisance than hazardous. The amount of fume generated from a furnace is a function of its efficiency. Since the early 1980s, furnace performance has improved considerably and consequently the amount of fume produced from a furnace has been reduced by a factor of two. In the late 1970s, the silica fume collected as a by-product presented a significant disposal problem. As of the mid-1990s, the silica fume is used as an additive for the production of cement and is in such demand that a silicon plant is known to operate for maximizing fume production and for silicon as a by-product.

A 75% ferrosilicon furnace operating at 90% silicon recovery would give 0.24 kg of by-product silica fume per kg of silicon produced. A silicon metal furnace operating at 85% recovery (typical of the industry) would give 0.38 kg of fume silica per kg silicon produced.

The principal health hazard that may be associated with silicon and silicon alloys is caused by the crystalline form of the oxide, ie, quartz, used as a raw material. Silica in its crystalline form is the chief cause of disabling pulmonary fibrosis, such as silicosis. Over a period of years, the breathing of air containing excessive amounts of crystalline silica can cause shortness of breath (30).

In the United States, amendments to the Clean Air Act in November 1990 limited the amount of sulfur dioxide emissions that coal-based power utilities could produce. The cost of compliance incurred by the utilities was expected to be passed along to the power consumers. The U.S. Bureau of Mines estimated that the requirements to limit sulfur dioxide emissions would increase the operational cost of certain silicon producers by up to \$0.02/kg (31).

Silicon and ferrosilicons by themselves are nontoxic. Some ferrosilicons may produce phosphorus hydride on coming in contact with water. In October 1993, the U.S. Department of Transportation classified ferrosilicon, with 30 and 90% silicon by weight, as "Dangerous When Wet." This is defined as "a material that, by contact with water, is liable to become spontaneously flammable or to give off flammable or toxic gas at a rate greater than one liter per kilogram of the material, per hour, when tested in accordance with paragraph 4 of Appendix E to this part" (32). This regulation has had significant impact on transportation of ferrosilicon alloys.

Following enactment of this regulation, a study was sponsored by Hickman, Williams & Company to evaluate the impact of gases evolved by the contact of water with 50%, 75%, and magnesium ferrosilicon. The study concluded that "commercial ferrosilicon does evolve phosphorus–hydride when

in contact with water. However, the quantity of gas evolved is substantially lower than the amount necessary to deem the reaction hazardous" (33).
Hickman, Williams and Company has filed an application for exemption with the U.S. Department of Transportation Administrator for Hazardous Materials Safety regarding 50 and 75% ferrosilicon, and magnesium ferrosilicon. At this writing (ca 1996), a decision on this application is still pending.

BIBLIOGRAPHY

"Silicon and Silicon Alloys" in *ECT* 1st ed., Vol. 12, pp. 360–365, by J. C. Vignos, Ohio Ferro-Alloys Corp.; "Metallurgical Silicon and Silicides" in *ECT* 2nd ed., Vol. 18, pp. 125–132, by E. G. Simms, Ohio Ferro-Alloys Corp.; "Silicon and Silicon Alloys–Metallurgical" in *ECT* 3rd ed., Vol. 20, pp. 846–854, by R. F. Silver, Elkem Metals Co.

1. H. Moissan, *The Electric Furnace*, 2nd ed., trans. by V. Lenker, Chemical Publishing Co., Easton, Pa., 1920.
2. M. Hansen, *Constitution of Binary Alloys*, 2nd ed., McGraw-Hill Book Co., Inc., New York, 1958.
3. F. A. Shunk, *Constitution of Binary Alloys*, 2nd Suppl., McGraw-Hill Book Co., Inc., New York, 1969.
4. B. Skei, in *Electric Furnace Conference Proceedings*, Vol. 48, New Orleans, La., 1990, p. 225.
5. J. Kr. Tuset, *Principles of Silicon Refining*, International Seminar on Refining and Alloying of Liquid Aluminum and Ferro-Alloys, Trondheim, Norway, Aug. 26, 1985.
6. J. Kr. Tuset, *Principles of Silicon Refining*, Silicon for Chemical Industry, Geiranger, Norway, June 1992, p. 1.
7. **U.S. Pat. 4,575,856** (Mar. 11, 1986), J. A. Persson.
8. R. F. Silver, in *Electric Furnace Conference Proceedings*, Vol. 41, Detroit, Mich., 1983, p. 175.
9. R. F. Silver, in *Electric Furnace Conference Proceedings*, Vol. 52, Nashville, Tenn., 1994, p. 361.
10. M. D. Young and R. F. Silver, in *Electric Furnace Conference Proceedings*, Vol. 51, Washington, D.C., 1993, p. 165.
11. **U.S. Pat. 4,798,659** (Jan. 17, 1989), V. D. Dosaj, D. H. Filsinger, and J. E. Trunzo.
12. V. D. Dosaj, J. B. May, and A. N. Arvidson, *Direct Current, Closed Furnace Silicon Technology*, Silicon for Chemical Industry II, Loen, Norway, June 1994, p. 9.
13. D. W. Ksinsik, *Silicon Metal Production in an Open DC Arc Furnace*, in Ref. 12, p. 25.
14. H. R. Larson and J. H. Welborn, *Revitalize the U.S. Silicon/Ferrosilicon Industry Through Energy Efficient Technology*, DOE Final Report No. DOE/AL/94598-1 (DE95010689), U.S. Dept. of Energy, Washington, D.C., Feb. 1995.
15. L. Nygaard and H. Brekken, *Production of Water Granulated Silicon*, in Ref. 12, p. 61.
16. B. Pachaly, *Process Development in the MCS-Production: For Example Water Granulated Silicon*, in Ref. 12, p. 55.
17. J. P. deLinde, *Silicon Metal; An Era of Growth and Prosperity?* Silicon for Silicones Conference III, Sandefjord, Norway, June 1996, p. 337.
18. J. B. May, *Silicon Metal Market*, Canadian Industrial Minerals Conference, Vancouver, B.C., Canada, Oct. 1995.
19. J. P. deLinde, *The Outlook for Silicon Metal*, in Ref. 12, p. 257.
20. L. D. Cunningham, *Silicon Annual Review*, U.S. Dept. of Interior, Bureau of Mines, Washington, D.C., 1991.
21. L. D. Cunningham, *Silicon Annual Review*, U.S. Dept. of Interior, Bureau of Mines, Washington, D.C., 1993.
22. L. D. Cunningham, *Silicon Annual Review*, U.S. Dept. of Interior, Bureau of Mines, Washington, D.C., 1994; "Monthly Price Averages," *Platt's Metals Week*, McGraw-Hill Book Co., Inc., New York, Apr. 1996.
23. R. Keiffer and F. Benesovsky, *Hartstoffé*, Springer-Verlag, Vienna, Austria, 1963, p. 455.
24. V. Dosaj, J. K. Tuset, and H. Holta, *Silicon and Ferrosilicon Refining Lecture Notes*, Iron and Steel Society Continuing Education Course, Nov. 1994.
25. D. N. Matter, *Silicon: Its Alloys and Their Use*, Ohio Ferro-Alloys Corp., Canton, Ohio, 1961.
26. *Pulverized Silvery Pig Iron*, Foote Mineral Co., Exton, Pa., 1980.
27. J. P. deLinde, "Ferrosilicon Future Production Trends," *10th International Ferroalloys Conference*, Barcelona, Spain, Nov. 1994.
28. L. D. Cunningham, *Silicon Annual Review*, U.S. Dept. of Interior, Bureau of Mines, Washington, D.C., 1992.
29. *Metals Handbook*, Vol. 1, 8th ed., American Society for Metals, Metals Park, Ohio, 1961.
30. N. I. Sax, *Dangerous Properties of Industrial Minerals*, 3rd ed., Reinhold Book Corp., New York, 1968.
31. L. D. Cunningham, *Silicon Annual Review*, U.S. Dept. of Interior, Bureau of Mines, Washington, D.C., 1990.
32. *Code of Federal Regulations*, Title 49, Parts 100–177, *Transportation*, U.S. Government Printing Office, Washington, D.C., 1993.
33. D. H. Gelwicks, L. A. Badanjek, C. L. Dodson, H. P. Jun, D. T. Rowe, and G. D. Rowe, *Electric Furnace Conference Proceedings*, Vol. 53, Orlando, Fla., 1995.

General References

Product data literature, Elkem Metals Co., Pittsburgh, Pa., 1996.
Product data literature, Globe Metallurgical Sales Inc., Cleveland, Ohio, 1996.
Product profile literature, Hickman, Williams and Company, Lakewood, Ohio, 1996.

Vishu Dosaj
Dow Corning Corporation